

Johnizer

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 15:18:20 ; Search time 73.49 Seconds
(without alignments)
21.160 Million cell updates/sec

Title: US-09-647-522-1
Perfect score: 72
Sequence: 1 GEIQTGRDVRGQAT 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	72	100.0	14	20	AAV33648	C. rastoni hemol
2	72	100.0	450	20	AAV33651	C. rastoni hemol
3	39	54.2	300	22	ABB69651	Drosophila melano
4	38	52.8	59	22	AAU05446	Propionibacterium
5	38	52.8	257	22	ABG06177	Novel human diagn
6	38	52.8	258	22	ABG05680	Novel human diagn
7	38	52.8	283	21	AA854021	Human pancreatic
8	38	52.8	913	22	AA876594	Corynebacterium 9
9	38	52.8	1997	21	AA819774	Human protein tyro
10	38	52.8	1997	22	AA878821	Human protein SED
11	38	52.8	2002	22	AA879805	Human protein SED

12	38	52.8	2993	22	AA62485	C glutamincum prote
13	37	51.4	122	22	AA046616	Propionibacterium
14	37	51.4	144	22	AA062661	Propionibacterium
15	37	51.4	250	22	AA690023	C glutamincum prote
16	37	51.4	250	22	AA680049	Corynebacterium q1
17	37	51.4	280	19	AA72879	Helicobacter pylori
18	37	51.4	339	19	AA698829	H. pylori GHP 148
19	37	51.4	359	22	AB627128	Novel human diagno
20	37	51.4	457	21	AA508875	Murine NN3 proteol
21	37	51.4	505	20	AA504839	Mycobacterium spec
22	37	51.4	876	22	AB600217	Novel human diagno
23	37	51.4	931	22	AA685010	Shrimp white spot
24	36	50.0	33	22	AB630783	Peptide #3434 enco
25	36	50.0	53	22	AB635960	Peptide #3466 enco
26	36	50.0	53	22	AB621370	Protein #3369 enco
27	36	50.0	53	22	AA656761	Human brain expres
28	36	50.0	53	22	AA656142	Human bone marrow
29	36	50.0	53	22	AA616974	Peptide #3408 enco
30	36	50.0	53	22	AA629463	Peptide #3500 enco
31	36	50.0	53	22	AA604674	Peptide #3356 enco
32	36	50.0	67	22	AB631708	Peptide #4359 enco
33	36	50.0	67	22	AB636928	Peptide #4434 enco
34	36	50.0	67	22	AB622248	Protein #4247 enco
35	36	50.0	67	22	AA657667	Human brain expres
36	36	50.0	67	22	AA670074	Human bone marrow
37	36	50.0	67	22	AA617900	Peptide #4334 enco
38	36	50.0	67	22	AA630410	Peptide #4447 enco
39	36	50.0	67	22	AA605553	Peptide #4235 enco
40	36	50.0	206	22	AB627693	Human peptide #344
41	36	50.0	206	22	AB632863	Peptide #369 enco
42	36	50.0	206	22	AB618345	Protein #344 enco
43	36	50.0	206	22	AA653666	Human brain expres
44	36	50.0	206	22	AA666049	Human bone marrow
45	36	50.0	206	22	AA613918	Peptide #352 enco

ALIGNMENTS

RESULT 1	
AA193648	
AA193648 standard; peptide: 14 AA.	
XX	
AC	AA193648;
XX	
DT	06-JAN-2000 (first entry)
XX	
DE	C. rastoni hemolytic protein derived peptide 1.
XX	
KW	Hemolytic protein; blood platelet agglutination; drug development;
KW	treatment; sting; jellyfish; pharmaceutical; pesticide.
XX	
OS	Carybdea rastoni.
XX	
PN	W09950294-A1.
XX	
PD	07-OCT-1999.
XX	
PF	30-MAR-1999; 99WO-JP01607.
XX	
PR	01-APR-1998; 98JP-0088569.
XX	
PA	(SUNR) SUNTORY LTD.
XX	
PI	Nagai H, Nakajima T;
XX	
DR	WPI: 1999-580740/49.
XX	
PT	Protein with hemolytic activity, useful for drugs treating jelly fish
PT	stings, pharmaceuticals with blood platelet agglutination activity,
PT	pesticides by use of the hemolytic activity, and study of the hemolytic
PT	mechanism
XX	

PS Claim 5; Page 22; 32pp; Japanese.
 CC This invention describes a novel protein which has hemolytic activity,
 CC blood platelet agglutination activity and a molecular weight of about
 CC 50,000 Da (by SDS-PAGE). The protein, homologs and analogs of the
 CC invention can be used as a novel approach to developing drugs useful for
 CC the treatment of jelly fish stings, pharmaceuticals with blood platelet
 CC agglutination activity, pesticides by use of the hemolytic activity, and
 CC in the study of the hemolytic mechanism. AAY33648-Y33650 represent
 CC fragments of the hemolytic protein described in the invention
 XX
 SQ Sequence 14 AA;
 Query Match 100.0%; Score 72; DB 20; Length 14;
 Best Local Similarity 100.0%; Pred. No. 6.2e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GEIQTKPPRVGQAT 14
 |||||
 Db 1 geiqtkpdrvqgat 14
 RESULT 2
 AAY33651
 ID AAY33651 standard; Protein; 450 AA.
 XX
 AC AAY33651;
 DT 06-JAN-2000 (first entry)
 XX
 NE C. rastonii hemolytic protein.
 KW *Hemolytic protein; blood platelet agglutination; drug development;
 KW treatment; sting; jellyfish; pharmaceutical; pesticide.
 OS Carybdea rastonii.
 PN WO9950294-A1.
 XX
 PD 07-OCT-1999.
 XX
 PF 30-MAR-1999; 99WO-JP01607.
 XX
 PR 01-APR-1998; 98JP-0088569;
 XX
 PA (SUNR) SUNTORY LTD.
 XX
 PI Nagai H, Nakajima T;
 XX
 DR WPI; 1999-580740/49.
 DR N-PSDB; AA23610.
 XX
 PT Protein with hemolytic activity, useful for drugs treating jelly fish
 PT stings; pharmaceuticals with blood platelet agglutination activity;
 PT pesticides by use of the hemolytic activity, and study of the hemolytic
 PT mechanism
 XX
 PS Claim 4; Page 26-27; 32pp; Japanese.
 CC This invention describes a novel protein which has hemolytic activity,
 CC blood platelet agglutination activity and a molecular weight of about
 CC 50,000 Da (by SDS-PAGE). The protein, homologs and analogs of the
 CC invention can be used as a novel approach to developing drugs useful for
 CC the treatment of jelly fish stings, pharmaceuticals with blood platelet
 CC agglutination activity, pesticides by use of the hemolytic activity, and
 CC in the study of the hemolytic mechanism. This sequence represents the
 CC hemolytic protein described in the invention
 XX
 SQ Sequence 450 AA;
 Query Match 100.0%; Score 72; DB 20; Length 450;

Best Local Similarity 100.0%; Pred. No. 0.00018;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GEIQTKPPRVGQAT 14
 |||||
 Db 56 geiqtkpdrvqgat 69
 RESULT 3
 ABB69651
 ID ABB69651 standard; Protein; 300 AA.
 XX
 AC ABB69651;
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 35745.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PE 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL13754.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 35745; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (ABB5737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 300 AA;
 Query Match 54.2%; Score 39; DB 22; Length 300;
 Best Local Similarity 61.5%; Pred. No. 72;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 Oy 2 EIQTKPDRVGOAT 14
 ::||| | ||| |
 Db 123 kmqtkvdavqgit 135
 RESULT 4
 AA055446
 ID AA055446 standard; Protein; 59 AA.
 XX

AC AAU55446;
XX
XX 27-FEB-2002 (first entry)
XX
XX
DE Propionibacterium acnes immunogenic protein #16342.
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
XX WO200181361-A2.
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
XX 02-JUN-2000; 2000US-208841P.
XX 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
XX WPI: 2001-616774/71.
XX N-PSDB: AAS59569.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris -
XX
XX Example 1: SEQ ID No 16641: 1069pp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX CC polypeptides. The proteins and their associated DNA sequences are used in
XX CC the treatment, prevention and diagnosis of medical conditions caused by
XX CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX CC P. acnes is also involved in infections of bone, joints and the central
XX CC nervous system, however it is particularly involved in the inflammatory
XX CC lesions associated with acne vulgaris. A method for detecting the
XX CC presence or absence of P. acnes in a patient comprises contacting a
XX CC sample with a binding agent that binds to the proteins of the invention
XX CC and determining the amount of bound protein in the sample. The
XX CC polypeptides may be used as antigens in the production of antibodies
XX CC specific for P. acnes proteins. These antibodies can be used to
XX CC downregulate expression and activity of P. acnes polypeptides and
XX CC therefore treat P. acnes infections. The antibodies may also be used as
XX CC diagnostic agents for determining P. acnes presence, for example, by
XX CC enzyme linked immunosorbent assay (ELISA).
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 59 AA:
SQ

Query Match 52.8%; Score 38; DB 22; Length 59;
Best Local Similarity 63.6%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
OY 4 QTRPDRVGQAT 14
Db 16 qtrpdrvgart 26

RESULT 5
ABG06177
ID ABG06177 standard; Protein: 257 AA.

XX
XX ABG06177;
XX
XX 13-FEB-2002 (first entry)
XX
XX
DE Novel human diagnostic protein #6168.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX N-PSDB: AAS70364.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX
XX Claim 20; SEQ ID No 36536; 1033p; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping; and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC and responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 257 AA:
SQ

Query Match 52.8%; Score 38; DB 22; Length 257;
Best Local Similarity 53.8%; Pred. No. 92;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
OY 2 EIORPDRVGQAT 14
Db 19 ktrpdrvgart 31

RESULT 6
ABG05880
ID ABG05880 standard; Protein: 258 AA.

AC ABG05880;
 XX
 DT 13-FEB-2002 (first entry)
 DE Novel human diagnostic protein #5871.
 DE
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001MO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEO INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR N-PSDB; AAS70067.
 XX
 DR WPI: 2001-639362/73.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 PT
 PS Claim 20; SEQ ID NO 36239; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 258 AA;

Query Match 52.8%; Score 38; DB 22; Length 258;
 Best Local Similarity 53.8%; Pred. No. 92;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 EIQTKEPRVGOAT 14
 : : | | | | | |
 Db 19 KIRVKPDRITGVLT 31

RESULT 7
 AAB54021
 ID AAB54021 standard; Protein; 283 AA.
 XX
 AC AAB54021;

XX
 DT 09-MAR-2001 (first entry)
 DE Human pancreatic cancer antigen protein sequence SEQ ID NO:473.
 DE
 XX Human: pancreas; pancreatic cancer; pancreatic cancer antigen;
 KM detection; diagnosis; identification; cytostatic; neuroprotective;
 KM neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KM antinflammatory; cardiast; gene therapy; chromosome mapping;
 KM linkage analysis; tissue identification; tissue typing; forensic;
 KM neural; immune system; muscular; reproductive; gastrointestinal;
 KM pulmonary; cardiovascular; renal; proliferative.
 XX
 OS Homo sapiens.
 XX
 PN WO200055320-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000MO-US05989.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-579444/54.
 DR N-PSDB; AAC98786.
 XX
 PT New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition -
 PT
 PS Claim 11; Page 909-910; 1379pp; English.
 XX
 CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, neotropic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiast and antinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 283 AA;

Query Match 52.8%; Score 38; DB 21; Length 283;
 Best Local Similarity 53.8%; Pred. No. 1e+02;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 EIQTKEPRVGOAT 14
 : : | | | | | |
 Db 47 KIRVKPDRITGVLT 59

RESULT 8
 AAB76594

ID AAB76594 standard; Protein: 913 AA.
XX
AC AAB76594;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein SEQ ID NO:170.
XX
KM Corynebacterium glutamicum; Corynebacterium lactofermentum; MCT;
KM membrane construction and membrane transport protein; petroleum spill;
KM hydrocarbon degradation; gram positive aerobic bacterium; marker;
KM identification; microorganism; fine chemical production; transformation;
KM genome mapping; genetic engineering.
XX
OS Corynebacterium glutamicum.
XX
PN MO200100805-A2.
XX
PD 04-JAN-2001.
XX
PE 23-JUN-2000; 2000WO-IB00926.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031563.
PR 09-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032190.
PR 09-JUL-1999; 99DE-1032191.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032212.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032927.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040830.
PR 27-AUG-1999; 99DE-1040831.
PR 27-AUG-1999; 99DE-1040832.
PR 27-AUG-1999; 99DE-1040833.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041395.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042088.
XX
PA (BADI) BASF AG.
XX
PI Pompeius M, Kroegeer B, Schroeder H, Zeidler O, Haberhauer G;
XX
DR WPI: 2001-071486/08.
DR N-PSDB; AAF67827.
XX
PT Corynebacterium glutamicum nucleic acids encoding membrane construction
PT and membrane transport proteins or their portions, useful for typing or
PT identifying C. glutamicum or related bacteria, and as markers for
PT transformation -
XX
PS Claim 20; Page 404-406; 111pp; English.
XX
CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane

CC construction and membrane transport (MCT) proteins given in AAB76510 to
CC AAB76847. The MCT nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points
CC for mapping C. glutamicum genome, and as markers for transformation.
CC AAF68082 and AAF68082 represent sequencing primers which are used in an
CC example from the present invention.
XX
SQ Sequence 913 AA;
XX
Query Match 52.8%; Score 38; DB 22; Length 913;
Best Local Similarity 58.3%; Pred. No. 3.2e+02;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 3 IOTKPRVGOAT 14
Db 263 vqtpdtagqvt 274
XX
RESULT 9
AAB19774
ID AAB19774 standard; Protein: 1997 AA.
XX
AC AAB19774;
XX
DT 19-FEB-2001 (first entry)
XX
DE Human protein tyrosine phosphatase HPTP-beta.
XX
XX Protein tyrosine phosphatase; HPTP-beta; human; Tie-2;
KM receptor-type tyrosine kinase; antiangiogenic; antitumor;
KM antineoplastic; tumour; metastasis; angiogenesis; therapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 1622..1641
FT /note="transmembrane domain"
FT Domain 1722..1967
FT /note="catalytic domain"
XX
PN EP1046715-A1.
XX
PD 25-OCT-2000.
XX
PE 23-APR-1999; 99EP-0108074.
XX
PR 23-APR-1999; 99EP-0108074.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Fachlinger G, Ritsau B, Deutsch U;
XX
DR WPI: 2000-648932/63.
DR N-PSDB; AAA88866.
XX
PT Monitoring or modulating Tie-2 tyrosine kinase activity, useful e.g.
PT for regulating tumor growth, using vascular-endothelial protein
PT tyrosine phosphatase -
XX
PS Disclosure; Page 21-27; 60pp; English.
XX
CC The present sequence is that of human protein tyrosine phosphatase
CC HPTP-beta, a member of subclass III receptor type PTPs, bearing
CC fibronectin type III-like repeats in the extracellular domain and a
CC single catalytic domain in the cytoplasmic tail. HPTP-beta is a
CC vascular-endothelial protein tyrosine phosphatase (VE-PTP) that
CC specifically interacts with receptor-type tyrosine kinase Tie-2,
CC modulating its tyrosine phosphorylation. Tie-2 is involved in
CC angiogenic processes, the formation of blood vessels during

CC embryonal development, wound healing and in pathological processes
CC such as tumour development. VE-pros such as HRP-beta or its
CC catalytic domain, nucleic acids and ligands can be used to monitor,
CC stimulate or repress Tie-2 activity for the purpose of monitoring
CC or modulating angiogenesis, inducing or inhibiting vascular growth
CC or remodelling and blood vessel maturation, and inhibiting tumour
CC growth or metastasis.

XX
XX
SQ Sequence 1997 AA;

Query Match

Best Local Similarity 52.8%; Score 38; DB 21; Length 1997;
Pred. No. 6.8e+02;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEIOTKPDYV 10

Db 1349 gsvrtpkpk1 1358

RESULT 10

AAM78821

ID AAM78821 standard; Protein; 1997 AA.

XX
AC AAM78821;

DT 06-NOV-2001 (first entry)

DE - Human protein SEQ ID NO 1483.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

XX WO200157190-A2.

PD 09-AUG-2001.

PF 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI: 2001-476283/51.

DR N-PSDB; AAK51954.

XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -

XX Claim 20; Page 3761-3764; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and
CC activating/inhibiting activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.

XX
XX
SQ Sequence 1997 AA;

Query Match

Best Local Similarity 52.8%; Score 38; DB 22; Length 1997;
Pred. No. 6.8e+02;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 GEIOTKPDYV 10

Db 1349 gsvrtpkpk1 1358

RESULT 11

AAM79805

ID AAM79805 standard; Protein; 2002 AA.

XX
AC AAM79805;

DT 06-NOV-2001 (first entry)

DE Human protein SEQ ID NO 3451.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

XX WO200157190-A2.

PD 09-AUG-2001.

PF 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI: 2001-476283/51.

DR N-PSDB; AAK52938.

XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -

XX Claim 20; Page 340-341; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAW00020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX

SO Sequence 2002 AA:

Query Match

Best Local Similarity 52.8%; Score 38; DB 22; Length 2002;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEIOTKPDV 10

DB 1354 gsvrtkpkdkl 1363

RESULT 12

AAAG92485 standard; Protein; 2993 AA.

AC AAG92485;

DT 26-SEP-2001 (first entry)

DE C glutamylcysteine protein fragment SEQ ID NO: 6239.

KW Corynebacterium; amino acid synthesis; vitamin; saccharide;

KM organic acid synthesis.

OS Corynebacterium glutamicum.

PN EPI108790-A2.

PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

XX N-PSDB: AAH67704.

PS Claim 17, SEQ ID NO: 6239; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein

CC sequences from the Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of corynebacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived

CC from Corynebacterium, and identifying a homologue of a gene derived

CC from corynebacterium. Corynebacterium bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a protein described

CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

XX Sequence 2993 AA:

Query Match

Best Local Similarity 52.8%; Score 38; DB 22; Length 2993;

Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 IORFDRVGOAT 14

DB 2343 vqltpdfagqvt 2354

RESULT 13

AAU46616 standard; Protein; 122 AA.

AC AAU46616;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #7512.

KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;

KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;

KM dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI) CORIAXA CORP.

XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI: 2001-616774/71.

XX N-PSDB: AASS9534.

PS Example 1, SEQ ID No 7811; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic

CC polypeptides. The proteins and their associated DNA sequences are used in

CC the treatment, prevention and diagnosis of medical conditions caused by

CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,

CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.

CC P. acnes is also involved in infections of bone, joints and the central

CC nervous system, however it is particularly involved in the inflammatory

CC lesions associated with acne vulgaris. A method for detecting the

CC presence or absence of P. acnes in a patient comprises contacting a

CC sample with a binding agent that binds to the proteins of the invention

CC and determining the amount of bound protein in the sample. The

CC polypeptides may be used as antigens in the production of antibodies

CC specific for P. acnes proteins. These antibodies can be used to

CC downregulate expression and activity of P. acnes polypeptides and

CC therefore treat P. acnes infections. The antibodies may also be used as

CC diagnostic agents for determining P. acnes presence, for example, by

CC enzyme linked immunosorbent assay (ELISA).

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 122 AA;

Query Match 51.4%; Score 37; DB 22; Length 122;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 PDRVGA 13
| | | | |
Db 52 pdrvga 58

RESULT 14

AAU62661
ID AAU62661 standard; Protein: 144 AA.

XX
AC AAU62661;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #23557.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

OS WO200181581-A2.

PN 01-NOV-2001.

PD 20-APR-2001; 2001WO-US12865.

PF 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

PA Skeiky YAM, Persing DH, Mitcham Jr, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI: 2001-616774/71.

DR N-PSDB; AAS59628.

XX Propionibacterium acnes polypeptides and nucleic acids useful for

PT vaccinating against and diagnosing infections, especially useful for

PT treating acne vulgaris -

PT Example 1; SEQ ID NO 23856; 1069pp; English.

PS
XX

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic

CC polypeptides. The proteins and their associated DNA sequences are used in

CC the treatment, prevention and diagnosis of medical conditions caused by

CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,

CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.

CC P. acnes is also involved in infections of bone, joints and the central

CC nervous system, however it is particularly involved in the inflammatory

CC lesions associated with acne vulgaris. A method for detecting the

CC presence or absence of P. acnes in a patient comprises contacting a

CC sample with a binding agent that binds to the proteins of the invention

CC and determining the amount of bound protein in the sample. The

CC polypeptides may be used as antigens in the production of antibodies

CC specific for P. acnes proteins. These antibodies can be used to

CC downregulate expression and activity of P. acnes polypeptides and

CC therefore treat P. acnes infections. The antibodies may also be used as

CC diagnostic agents for determining P. acnes presence, for example, by

CC enzyme linked immunosorbent assay (ELISA).

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 144 AA;

Query Match 51.4%; Score 37; DB 22; Length 144;
Best Local Similarity 54.5%; Pred. No. 78;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 2 EIQTKPDRVGA 12
| | | | |
Db 126 elevpqrvgq 136

RESULT 15

AAG90023
ID AAG90023 standard; Protein: 250 AA.

XX
AC AAG90023;

DT 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 3777.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.

XX Corynebacterium glutamicum.

OS EPI108790-A2.

PN 20-JUN-2001.

PD 18-DEC-2000; 2000EP-0127688.

PF 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

DR N-PSDB; AAH65242.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

PT Claim 17; SEQ ID NO: 3777; 246pp + Sequence Listing; English.

PS
XX

XX The present invention provides a number of nucleotide and protein

CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of coryneform bacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived

CC from Coryneform bacterium, and identifying a homologue of a gene derived

CC from coryneform bacterium. Coryneform bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a protein described

CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

CC European Patent Office.

XX
SQ Sequence 250 AA;

Query Match 51.4%; Score 37; DB 22; Length 250;

Best Local Similarity 54.5%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 GEIQKPPDRVG 11
||:| | | |
Db 128 gevettptcrag 138

Search completed: June 10, 2002, 15:21:06
Job time: 166 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 15:19:00 ; Search time 35.61 seconds

(without alignments)
37.777 Million cell updates/sec

Title: us-09-647-522-1

Perfect score: 72

Sequence: 1 GEIOTKPDVVGQAT 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	72	100.0	450	2 JC7371	toxin - jellyfish
2	44	61.1	1219	2 T14578	nucleoporin Nup153
3	40	55.6	997	2 T28872	hypothetical prote
4	40	55.6	2145	2 JC4747	adenylate cyclase
5	38	52.8	255	1 S32482	electron transfer
6	38	52.8	317	2 AH0502	transaldolase B [l
7	38	52.8	368	2 D81256	probable ATP/GTP-b
8	38	52.8	476	2 S76518	hypothetical prote
9	38	52.8	623	2 F69093	Rhase I inhibitor
10	38	52.8	735	2 AE2169	cation transport A
11	38	52.8	1360	2 T06689	zinc finger protei
12	38	52.8	1997	1 S12050	protein-tyrosine-p
13	37	51.4	214	1 B69350	hypothetical prote
14	37	51.4	241	2 T50406	hypothetical prote
15	37	51.4	304	2 AD1029	probable membrane
16	37	51.4	311	2 A84321	hypothetical prote
17	37	51.4	339	2 F64556	aliphatic amidase
18	37	51.4	339	2 B71951	RNA-directed RNA p
19	37	51.4	500	2 A36388	transposase (clone
20	37	51.4	601	2 T36323	hypothetical prote
21	36.5	50.7	151	2 S37003	ribosomal protein
22	36	50.0	66	2 PNO644	probable mitochond
23	36	50.0	116	1 R5HSH9	probable membrane
24	36	50.0	192	2 T39418	o-pyrocatechuate d
25	36	50.0	206	2 T34961	hypothetical prote
26	36	50.0	292	2 S65362	hypothetical prote
27	36	50.0	293	2 T33825	epoxide hydrolase
28	36	50.0	303	2 AG1178	
29	36	50.0	455	1 A29939	

30	36	50.0	479	2 T15065	hypothetical prote
31	36	50.0	569	2 D69374	type II secretion
32	36	50.0	625	2 T37719	thymidylate syntha
33	36	50.0	628	2 A81999	glucose inhibited
34	36	50.0	631	2 F81227	glucose inhibited
35	36	50.0	634	2 F82079	probable 2',3'-cyc
36	36	50.0	714	2 A82366	hypothetical prote
37	36	50.0	727	2 AB0180	hypothetical prote
38	36	50.0	790	2 T30081	hypothetical prote
39	36	50.0	843	2 D96495	unknown protein [l
40	36	50.0	843	2 A87275	TonB-dependent rec
41	36	50.0	881	2 G96574	hypothetical prote
42	36	50.0	938	2 C84480	hypothetical prote
43	36	50.0	1171	2 T28701	probable polyketid
44	36	50.0	1410	2 T47137	hypothetical prote
45	36	50.0	2748	2 S57976	nuclear migration

ALIGNMENTS

RESULT 1
JC7371
toxin - jellyfish (Carybdea rastoni)
C:Species: Carybdea rastoni
C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 17-Nov-2000
C:Accession: JC7371; PC7094
R:Nagai, H.; Takawa, K.; Nakao, M.; Ito, E.; Miyake, M.; Noda, M.; Nakajima, T.
Biochem. Biophys. Res. Commun. 275, 582-588, 2000
A:Title: Novel proteinaceous toxins from the box jellyfish (sea wasp) Carybdea raston
A:Reference number: JC7371
A:Accession: JC7371
A:Molecule type: mRNA
A:Residues: 1-450 <NAG>
A:Cross-references: DDBJ:AB015878.
A:Accession: PC7094
A:Molecule type: protein
A:Residues: 39-55;56-70;196-210;250-267;268-279;309-325;363-377;378-382 <NA2>
C:Comment: This protein, a member of bioactive protein, has hemolytic activity.
C:Keywords: hemolysis; Inflammation; toxin

Query Match 100.0%; Score 72; DB 2; Length 450;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEIOTKPDVVGQAT 14
Db 56 GEIOTKPDVVGQAT 69

RESULT 2
T14578
nucleoporin Nup153 homolog - African clawed frog (fragment)
C:Species: Xenopus laevis (African clawed frog)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C:Accession: T14578
R:Shah, S.; Tugendreich, S.; Forbes, D.J.
J. Cell Biol. 141, 31-49, 1998
A:Title: Major binding sites for the nuclear import receptor are the internal nucleop
A:Reference number: Z18147; MUID:58198465
A:Accession: T14578
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1219 <SHA>
A:Cross-references: EMBL:AF045567; NID:g3047239; PID:g3047240; PIDN:AC41273.1

Query Match 61.1%; Score 44; DB 2; Length 1219;
Best Local Similarity 57.1%; Pred. No. 11;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Qy 1 GEIOTKPDVVGQAT 14

Db 639 GSAQTKPDPTLSQST 652

RESULT 3

T28872

hypothetical protein R04E5.8 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000

C:Accession: T28872

R:Miller, N.

submitted to the EMBL Data Library, December 1995

A:Description: The sequence of *C. elegans* cosmid R04E5.

A:Reference number: Z20535

A:Accession: T28872

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-997 <ML>

A:Cross-references: EMBL:U41538; PIDN:AAC48181.1; GSPDB:GN00028; CESP:R04E5.8

A:Experimental source: strain Bristol N2; clone R04E5

C:Genetics:

A:Gene: CESP:R04E5.8

A:Map position: X

A:Introns: 216/3; 234/3; 255/2; 305/1; 669/3; 891/1; 986/2

Query Match

Best Local Similarity 55.6%; Score 40; DB 2; Length 997;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 4 QTKPDVGOA 13

Db 985 EARDRGOA 994

RESULT 4

JC4747

adenylate cyclase (EC 4.6.1.1) - *Podospora anserina*

N:Alternate names: 3',5'-Cyclic AMP synthetase; adenylylcyclase

C:Species: *Podospora anserina*

C:Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 21-Jul-2000

C:Accession: JC4747

R:Loubradou, G.; Begueret, J.; Turcq, B.

Gene 170, 119-123, 1996

A:Title: An additional copy of the adenylate cyclase-encoding gene relieves development

A:Reference number: JC4747; MUID:96200867

A:Accession: JC4747

A:Molecule type: DNA

A:Residues: 1-2145 <LOU>

A:Cross-references: GB:L43413; NID:g1261824; PIDN:AAB05642.1; PID:g1480208

C:Comment: This protein is produced by a mutation in a vegetative incompatibility-contro

C:Genetics:

A:Gene: PAC

A:Introns: 79/2; 319/3; 2091/3

C:Superfamily: yeast adenylate cyclase; leucine-rich alpha-2-glycoprotein repeat homolog

C:Keywords: leucine zipper; phosphorus-oxygen lyase

F:694-1372/Domain: leucine-rich #status predicted <LEU>

F:826-849/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>

F:850-872/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>

F:873-895/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>

F:896-919/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>

F:943-964/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>

F:1007-1029/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>

F:1259-1282/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>

F:1283-1306/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>

F:1748-1991/Domain: catalytic #status predicted <CAT>

F:1769-1854/Domain: yeast adenylate cyclase catalytic domain homology <YACC>

F:1994-1999,2125-2129/Domain: CAP-binding #status predicted <CAP>

F:2035-2063/Region: leucine zipper motif

Query Match

Best Local Similarity 55.6%; Score 40; DB 2; Length 2145;

Matches 66.7%; Pred. No. 1.1e+02;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 3 IOTKPDVGOA 14

Db 907 IOTLDPNVGOAT 918

RESULT 5

S32482

electron transfer flavoprotein beta chain - human

C:Species: *Homo sapiens* (man)

C:Date: 06-Jan-1995 #sequence_revision 12-Jul-1996 #text_change 11-Jun-1999

C:Accession: S32482

R:Finocchiaro, G.; Colombo, I.; Garavaglia, B.; Gellera, C.; Valdameri, G.; Garbuglio

Eur. J. Biochem. 213, 1003-1008, 1993

A:Title: cDNA cloning and mitochondrial import of the beta-subunit of the human elect

A:Reference number: S32482; MUID:93279298

A:Accession: S32482

A:Molecule type: mRNA

A:Residues: 1-255 <FIN>

A:Cross-references: EMBL:X71129; NID:g297901; PIDN:CA50441.1; PID:g297902

C:Comment: This protein does not have a cleavable transit peptide.

C:Genetics:

A:Gene: GDB:ETFB

A:Cross-references: GDB:119887; OMIM:130410

A:Map position: 19q13.4-19q13.4

C:Function: heterodimer of alpha (see PIR:A31998) and beta chains that binds one molec

C:Description: electron transfer from any of several dehydrogenases in the mitochondr

C:Superfamily: electron transfer flavoprotein beta chain

C:Keywords: electron transfer; FAD; flavoprotein; heterodimer; mitochondrial matrix

Query Match

Best Local Similarity 52.8%; Score 38; DB 1; Length 255;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 EIOKPDVGOAT 14

Db 19 KIRKVPRTGVT 31

RESULT 6

AH0502

transaldolase B (imported) - *Salmonella enterica* subsp. *enterica* serovar Typhi (stralC:Species: *Salmonella enterica* subsp. *enterica* serovar TyphiA:Note: this species has also been called *Salmonella typhi*

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001

C:Accession: AH0502

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moute, S.; O'Goara, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* se

A:Reference number: AH0502; PMID:11677608

A:Accession: AH0502

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-317 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD01160.1; PID:g16501290; GSPDB:GN00176

C:Genetics:

A:Gene: SRY0007

C:Superfamily: human transaldolase

Query Match

Best Local Similarity 52.8%; Score 38; DB 2; Length 317;

Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

OY 1 GEIOTKPDVGOA 13

Db 267 GEVNAPEPRTTEA 279

```

RESULT 7
DB1256
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Sep-2000
C:Accession: DB1256
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kellley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barre
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912
A:Accession: DB1256
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-368 <PAR>
A:Cross-references: GB:AL139079; GB:AL111168; NID:g6968971; PIDN:CAB75594.1; PID:g696902
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: mmp; Cj1606c
C:Superfamily: conserved probable membrane protein Y1L003w

Query Match 52.8%; Score 38; DB 2; Length 368;
Best Local Similarity 58.3%; Pred. No. 38;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 GEIOTKPRVGO 12
DB 146 GEIOTKPRVGO 157

RESULT 8
S76518
C:Species: Synechocystis sp. (strain PCC 6803)
C:Accession: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
S.
A:Reference number: S74322; MUID:97061201
A:Accession: S76518
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-476 <KAN>
A:Cross-references: EMBL:D64002; GB:AB001339; NID:g1001612; PIDN:BA10364.1; PID:dl01101
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Start codon: GTG

Query Match 52.8%; Score 38; DB 2; Length 476;
Best Local Similarity 56.2%; Pred. No. 50;
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

OY 1 GEIOTKPR--DRVGOAT 14
DB 440 GEVRLPNDGRAGQAT 455

RESULT 9
F69093
C:Species: Methanobacterium thermoautotrophicum (strain Delta H)
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Feb-2001
C:Accession: F69093
R:Smith, D.R.; Doucette-Stamm, L.A.; Delouhery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Olun, D.; Spadefora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.

```

```

ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A:Reference number: A69000; MUID:98037514
A:Accession: F69093
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-623 <MTH>
A:Cross-references: GB:AE000927; GB:AE000666; NID:g2622822; PIDN:AAB86167.1; PID:g262
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1695
C:Superfamily: ribonuclease L inhibitor; ATP-binding cassette homology; ferredoxin 21
C:Keywords: ATP; nucleotide binding; P-loop
F:37-101/Domain: ferredoxin 21(4e-4s) homology <FER2>
F:119-318/Domain: ATP-binding cassette homology <ABC1>
F:137-144/Region: nucleotide-binding motif A (P-loop)
F:386-561/Domain: ATP-binding cassette homology <ABC2>
F:403-410/Region: nucleotide-binding motif A (P-loop)

Query Match 52.8%; Score 38; DB 2; Length 623;
Best Local Similarity 58.3%; Pred. No. 67;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 EIOTKPRVGOA 13
DB 356 EFVKPRPVGEA 367

RESULT 10
AE2169
C:Species: Anabaena sp.
C:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AE2169
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kurita, T.; Sasamoto, S.; Watanabe, A.; Irigun
Nakasaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2169
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-735 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW74607.1; PID:g17132002; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2908

Query Match 52.8%; Score 38; DB 2; Length 735;
Best Local Similarity 57.1%; Pred. No. 81;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 GEIOTKPRVGOAT 14
DB 297 GEIYIOAERVGTAT 310

RESULT 11
T06699
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Oct-1999
C:Accession: T06699
R:Queller, F.; Choisey, N.; Robert, C.; Brotter, P.; Wincker, P.; Catolico, L.; Art
submitted to the Protein Sequence Database, April 1999
A:Reference number: Z15793
A:Accession: T06699
A:Molecule type: DNA
A:Residues: 1-1360 <QUE>

```

A:Cross-references: EMBL:AL049659; GSPDB:GN00061; ATSP:T29H11.50
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Species: Homo sapiens (man)
C:Accession: S12050; S15818; S15819
R:Krieger, N.X.; Streuli, M.; Saito, H.
EMBO J. 9, 3241-3252, 1990
A:Title: Structural diversity and evolution of human receptor-like protein tyrosine phosphatase
A:Reference number: S12049; MUID:91006018
A:Accession: S12050
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1997 <RNU>
A:Cross-references: GB:X54131; NID:935787; PIDN:CA38066.1; PID:935788
R:de Vries, L.; Li, R.Y.; Ragab, A.; Ragab-Thomas, J.M.F.; Chap, H.
FEBS Lett. 282, 285-288, 1991
A:Title: Expression of a truncated protein-tyrosine phosphatase mRNA in human lung.
A:Reference number: S15818; MUID:91243813
A:Accession: S15818
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1872-1911, 'VHMYLQK' <VRI>
A:Accession: S15819
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1872-1997 <VR2>
C:Genetics:
A:Gene: GDB:PTPRB; PTPB
A:Cross-references: GDB:127352; OMIM:176882
A:Map position: 12q15-12q21
C:Superfamily: protein-tyrosine-phosphatase, receptor type beta; fibronectin type III rd
C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; receptor
F:122/Domain: signal sequence #status predicted <SIG>
F:23-1997/Product: protein-tyrosine-phosphatase, receptor type beta #status predicted <M
F:23-1625/Domain: extracellular #status predicted <EXT>
F:1626-1642/Domain: transmembrane #status predicted <TMN>
F:1643-1997/Domain: intracellular #status predicted <INT>
F:1727-1952/Domain: protein-tyrosine-phosphatase homology <PMP>
F:1904/Active site: Cys (phosphocysteine intermediate) #status predicted
F:1910/Binding site: substrate phosphate (Arg) #status predicted

Query Match 52.8%; Score 38; DB 2; Length 1360;
Best Local Similarity 41.7%; Pred. No. 1.6e+02;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Y 1 GEIOTKPDRCVQ 12
Db 1195 GSELEKPKKICK 1206

RESULT 12

S12050
Protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type beta precursor - human

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S12050; S15818; S15819
R:Krieger, N.X.; Streuli, M.; Saito, H.
EMBO J. 9, 3241-3252, 1990
A:Title: Structural diversity and evolution of human receptor-like protein tyrosine phosphatase
A:Reference number: S12049; MUID:91006018
A:Accession: S12050
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1997 <RNU>
A:Cross-references: GB:X54131; NID:935787; PIDN:CA38066.1; PID:935788
R:de Vries, L.; Li, R.Y.; Ragab, A.; Ragab-Thomas, J.M.F.; Chap, H.
FEBS Lett. 282, 285-288, 1991
A:Title: Expression of a truncated protein-tyrosine phosphatase mRNA in human lung.
A:Reference number: S15818; MUID:91243813
A:Accession: S15818
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1872-1911, 'VHMYLQK' <VRI>
A:Accession: S15819
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1872-1997 <VR2>
C:Genetics:
A:Gene: GDB:PTPRB; PTPB
A:Cross-references: GDB:127352; OMIM:176882
A:Map position: 12q15-12q21
C:Superfamily: protein-tyrosine-phosphatase, receptor type beta; fibronectin type III rd
C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; receptor
F:122/Domain: signal sequence #status predicted <SIG>
F:23-1997/Product: protein-tyrosine-phosphatase, receptor type beta #status predicted <M
F:23-1625/Domain: extracellular #status predicted <EXT>
F:1626-1642/Domain: transmembrane #status predicted <TMN>
F:1643-1997/Domain: intracellular #status predicted <INT>
F:1727-1952/Domain: protein-tyrosine-phosphatase homology <PMP>
F:1904/Active site: Cys (phosphocysteine intermediate) #status predicted
F:1910/Binding site: substrate phosphate (Arg) #status predicted

Query Match 52.8%; Score 38; DB 1; Length 1997;
Best Local Similarity 50.0%; Pred. No. 2.4e+02;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Y 1 GEIOTKPDRCVQ 10
Db 1349 GSVTRKPKDKI 1358

RESULT 13

B69350
hypothetical protein AF0802 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: B69350
R:Kienk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kittness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Atliach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Moese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69250; MUID:98049343
A:Accession: B69350
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-214 <RUE>
A:Cross-references: GB:AE001049; GB:AE000782; NID:92689372; PIDN:AA90445.1; PID:9264
C:Superfamily: hypothetical protein AF0802

Query Match 51.4%; Score 37; DB 1; Length 214;
Best Local Similarity 45.5%; Pred. No. 33;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Y 1 GEIOTKPDRCVQ 11
Db 80 GELEQPDKICG 90

RESULT 14

T50406
hypothetical protein SPBP4G3.03 [imported] - fission yeast (Schizosaccharomyces pombe)

C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: T50406
R:Rieger, M.; McDougall, R.C.; Rajandream, M.A.; Barrett, B.G.
submitted to the EMBL Data Library, January 2000
A:Reference number: Z25068
A:Accession: T50406
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-241 <RIE>
A:Cross-references: EMBL:AL137099; PIDN:CAB8658.1; GSPDB:GN00067; SPDB:SPBP4G3.03
C:Experimental source: strain 972h(-); clone pl p4G3
C:Genetics:
A:Gene: SPDB:SPBP4G3.03
A:Map position: 2

Query Match 51.4%; Score 37; DB 2; Length 241;
Best Local Similarity 75.0%; Pred. No. 37;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Y 1 GEIOTKPD 8
Db 91 GQIETKPD 98

RESULT 15

AD1029
probable membrane protein SRY4557 [imported] - Salmonella enterica subsp. enterica se

C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: AD1029
R:Perkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.W.; Dowd, L.; White, N.; Farr
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; PMID:11677608
A:Accession: AD1029
A:Status: preliminary

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: June 10, 2002, 15:21:45 ; Search time 18'65 Seconds
(Without alignments)
29.066 Million cell updates/sec

Title: US-09-647-522-1
Sequence: 1 GEIOTKPRVGOAT 14
Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	55.6	2145	1	CYAA_PODAN
2	38	52.8	255	1	ETFB_PODAN
3	38	52.8	773	1	YHGF_ECOLI
4	38	52.8	1997	1	PRPB_HUMAN
5	37	51.4	311	1	MCH_HUMAN
6	37	51.4	443	1	ZIC1_XENLA
7	37	51.4	500	1	RRPA_CVMJD
8	36	50.0	115	1	DHBD_HALMA
9	36	50.0	292	1	RIBB_ASPNG
10	36	50.0	340	1	Y825_STRCO
11	36	50.0	455	1	HYEP_HUMAN
12	36	50.0	1816	1	AF6_HUMAN
13	36	50.0	2748	1	NUM1_HUMAN
14	35	48.6	395	1	THB_PAROL
15	35	48.6	465	1	HYDH_ECOLI
16	35	48.6	480	1	RADA_MYCTU
17	35	48.6	584	1	YMB3_YEAST
18	35	48.6	630	1	GIDA_YEAST
19	35	48.6	641	1	MI42_PSEPU
20	35	48.6	658	1	SYM_MOUSE
21	35	48.6	700	1	ERG_YEAST
22	35	48.6	807	1	SUSI_HOYU
23	35	48.6	4488	1	RRPA_CVMJD
24	35	48.6	160	1	YCIL_AQUAE
25	34	47.2	357	1	KAPC_YEAST
26	34	47.2	415	1	PROA_XLFLA
27	34	47.2	446	1	BRNO_LACDL
28	34	47.2	447	1	ZIC1_HUMAN
29	34	47.2	447	1	ZIC1_MOUSE
30	34	47.2	447	1	ZIC1_HUMAN
31	34	47.2	447	1	ZIC1_MOUSE
32	34	47.2	447	1	ZIC1_MOUSE
33	34	47.2	447	1	ZIC1_MOUSE
34	34	47.2	447	1	ZIC1_MOUSE
35	34	47.2	447	1	ZIC1_MOUSE
36	34	47.2	447	1	ZIC1_MOUSE
37	34	47.2	447	1	ZIC1_MOUSE
38	34	47.2	447	1	ZIC1_MOUSE
39	34	47.2	447	1	ZIC1_MOUSE
40	34	47.2	447	1	ZIC1_MOUSE
41	34	47.2	447	1	ZIC1_MOUSE
42	34	47.2	447	1	ZIC1_MOUSE
43	34	47.2	447	1	ZIC1_MOUSE
44	34	47.2	447	1	ZIC1_MOUSE
45	34	47.2	447	1	ZIC1_MOUSE

ALIGNMENTS

RESULT	1	STANDARD	PRT	2145 AA
AC	CYAA_PODAN			
DF	01-NOV-1997 (Rel. 35, Created)			
DF	16-OCT-2001 (Rel. 40, Last sequence update)			
DE	Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (AdenylL)			
OS	Podospora anserina.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OX	NCBI_TaxID=5145;			
RP	SEQUENCE FROM N.A.			
RA	Medline-96200867; Pubmed-8621071;			
RT	Loubradou G., Begueret J., Turc B.;			
RT	An additional copy of the adenylate cyclase-encoding gene relieves			
RL	developmental defects produced by a mutation in a vegetative			
CC	gene 170:119-123(1996).			
CC	- FUNCTION: PLAYS ESSENTIAL ROLES IN REGULATION OF CELLULAR			
CC	CAMP.			
CC	- CATALYTIC ACTIVITY: ATP -> 3',5'-cyclic AMP + diphosphate.			
CC	- SIMILARITY: BELONGS TO ADENYLATE CYCLASE CLASS-3 FAMILY.			
CC	- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).			
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL: I43413; AAB05642.1;			
DR	InterPro: IPR001054; Guanylyl_cyclase.			
DR	InterPro: IPR001611; LRR.			
DR	InterPro: IPR003592; LRR.			
DR	InterPro: IPR003591; LRR_out.			
DR	InterPro: IPR001932; PP2C_domain.			
DR	Pfam: PR00211; guanylate_cyc. 2.			
DR	Pfam: PF00560; LRR. 15.			
DR	Pfam: PF00481; PP2C. 1.			
DR	PRINTS: PR00019; LEUCRICHPT.			
DR	SMART: SM00044; CYCC. 1.			
DR	SMART: SM00370; LRR. 5.			
DR	SMART: SM00369; LRR_TYF. 1.			
DR	SMART: SM00332; PP2C. 1.			
DR	SMART: SM00314; RA. 1.			
DR	PROSITE: PS50125; GUANYLATE_CYCLASES_2. 1.			
KW	LYASE; Repeat: Leucine-rich repeat; CAMP synthesis; Magnesium.			
FT	REPEAT 751 776 LRR 1.			

```

FT REPEAT 778 800 LRR 2.
FT REPEAT 801 824 LRR 3.
FT REPEAT 826 847 LRR 4.
FT REPEAT 848 870 LRR 5.
FT REPEAT 872 894 LRR 6.
FT REPEAT 895 917 LRR 7.
FT REPEAT 919 940 LRR 8.
FT REPEAT 941 965 LRR 9.
FT REPEAT 983 1005 LRR 10.
FT REPEAT 1026 1027 LRR 11.
FT REPEAT 1028 1051 LRR 12.
FT REPEAT 1053 1074 LRR 13.
FT REPEAT 1075 1097 LRR 14.
FT REPEAT 1099 1120 LRR 15.
FT REPEAT 1233 1256 LRR 16.
FT REPEAT 1257 1280 LRR 17.
FT REPEAT 1282 1304 LRR 18.
FT REPEAT 1305 1328 LRR 19.
FT REPEAT 1357 1380 LRR 20.
FT REPEAT 1394 1419 LRR 21.
FT REPEAT 1445 1710 PP2C-LIKE.
FT DOMAIN 1711 2145 CATALYTIC.
FT DOMAIN 36 41 POLY-SER.
SQ SEQUENCE 2145 AA; 237515 MW; 88E7E6E4AC0687D CRC64;

Query Match 55.6%; Score 40; DB 1; Length 2145;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 3 IQTKPRVGOAT 14
Db 907 IQSLPDVGOAT 918

RESULT 2
ID ETEFB_HUMAN STANDARD; PRT; 255 AA.
AC P38117;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Electon transfer flavoprotein beta-subunit (Beta-ETFB).
GN ETEFB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID:9606;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE-Fetal liver;
RC MEDLINE-93279298; PubMed-8504797;
RX Fmocchiaro G., Colombo I., Garavaglia B., Gellera C., Valdemeri G.,
RA Garbuglio N., Didonato S.;
RT "cdna cloning and mitochondrial import of the beta-subunit of the
RT human electron-transfer flavoprotein."
RL Eur. J. Biochem. 213:1003-1008(1993).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX MEDLINE-97121389; PubMed-8962055;
RA Roberts D.L., Freeman F.E., Kim J.J.;
RT "Three-dimensional structure of human electron transfer flavoprotein
RT to 2.1-A resolution."
RL Proc. Natl. Acad. Sci. U.S.A. 93:14355-14360(1996).
RN [3]
RP VARIANT GAIB GLN-164.
RX MEDLINE-94282034; PubMed-7912128;
RA Colombo I., Fmocchiaro G., Garavaglia B., Garbuglio N.,
RA Yamaguchi S., Freeman F., Berra B., Didonato S.;
RT "Mutations and polymorphisms of the gene encoding the beta-subunit of
RT the electron transfer flavoprotein in three patients with glutaric
RT acidemia type II."
RL Hum. Mol. Genet. 3:429-435(1994).

```

```

RN [4]
RP VARIANT MET-154.
RX MEDLINE-99286383; PubMed-10356313;
RA Bross P., Pedersen P., Winter V., Nyholm M., Johansen B.N.,
RA Olsen R.K., Corydon M.J., Andresen B.S., Elberg H., Kolvraa S.,
RA Gregersen N.;
RT alpha-chain (alpha-T171) displays decreased thermal stability and is
RT overrepresented in very-long-chain acyl-CoA dehydrogenase-deficient
RT patients with mild childhood presentation."
RL Mol. Genet. Metab. 67:138-147(1999).
SQ SEQUENCE 255 AA; 27843 MW; 47E6AEF50EB2C80 CRC64;

Query Match 52.8%; Score 38; DB 1; Length 255;
Best Local Similarity 53.8%; Pred. No. 11;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 EIOTKPRVGOAT 14
Db 19 KIRYKPRVGOAT 31

RESULT 3
ID YHGF_ECOLI STANDARD; PRT; 773 AA.
AC P46837; P76689;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 36, Last sequence update)

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

-1- SIMILARITY: BELONGS TO THE ETF BETA-SUBUNIT / FIXA FAMILY.
 OF LASTING LEHARGY OR SLOWLY PROGRESSIVE MYOPATHY.
 KIDNEYS. THE Milder FORM CAN BE CHARACTERIZED BY RECURRENT EPISODES
 DEATH. THE Milder FORM CAN BE CHARACTERIZED BY RECURRENT EPISODES
 ITS MOST SEVERE FORM IN INFANCY, WITH POLYCYSTIC AND DYSPLASTIC
 ISOBUTYRIC, 2-METHYL-BUTYRIC, AND ISOVALLERIC ACIDS. IT IS SEEN IN
 ONLY OF GLUTARIC ACID BUT ALSO OF LACTIC, ETHYLMALONIC, BUTYRIC,
 IIB (GALIB) (GLUTARICACIDURIA) WHICH RESULTS IN THE EXCRETION NOT
 -1- DISEASE: DEFECTS IN ETEFB ARE THE CAUSE OF GLUTARIC ACIDURIA TYPE
 AND PANCHEAS.
 -1- FUNCTION: THE ELECTRON TRANSFER FLAVOPROTEIN SERVES AS A SPECIFIC
 MOLECULAR ACCEPTOR FOR SEVERAL DEHYDROGENASES, INCLUDING FIVE ACYL-
 COA DEHYDROGENASES, GLUTARYL-COA AND SARCOSINE DEHYDROGENASE. IT
 TRANSFERS THE ELECTRONS TO THE MAIN MITOCHONDRIAL RESPIRATORY
 CHAIN VIA ETF-UBIQUINONE OXIDOREDUCTASE (ETF DEHYDROGENASE).
 -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
 -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
 -1- TISSUE SPECIFICITY: ABUNDANT IN LIVER, HEART AND SKELETAL MUSCLE.
 -1- TISSUE EXPRESSION IS SEEN IN THE BRAIN, PLACENTA, LUNG, KIDNEY
 A WEAK EXPRESSION IS SEEN IN THE BRAIN, PLACENTA, LUNG, KIDNEY
 AND PANCHEAS.

EMBL: X71129; CAA50441.1; NOT_ANNOTATED_CDS.
 DR EMBL: X76067; NOT_ANNOTATED_CDS.
 DR PIR: S32482; S32482.
 DR PDB: 1EFV; 03-DEC-97.
 DR MIM: 130410; -
 DR InterPro: IPR000049; ETF_beta.
 DR InterPro: P01012; ETF_beta: 1.
 DR Pfam: PF003528; ETF_beta: 1.
 DR PROSITE: PS01065; ETF_BETA: 1.
 DR Electon transport; Flavoprotein; FAD; Mitochondrion;
 KW Disease mutation; Glutaricaciduria; Polymorphism; 3D-structure.
 FT VARIANT 154 154 /FTID-VAR_008548.
 FT VARIANT 164 164 R->Q (IN GALIB).
 FT VARIANT 164 164 /FTID-VAR_002369.
 FT VARIANT 164 164 /FTID-VAR_002369.
 SQ SEQUENCE 255 AA; 27843 MW; 47E6AEF50EB2C80 CRC64;

RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE-Placenta;
RX	MEDLINE=91006018; Pubmed=2170109;
RA	Krueger N.X., Streuli M., Saito H.;
RT	"Structural diversity and evolution of human receptor-like protein tyrosine phosphatases";
RL	EMBO J. 9:3241-3252(1990).
CC	-I CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein lysosine + phosphate.
CC	-I SUBCELLULAR LOCATION: Type I membrane protein.
CC	-I SIMILARITY: CONTAINS 1 PROTEIN-TYROSINE PHOSPHATASE DOMAIN. -I SIMILARITY: CONTAINS 16 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as their content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isdb-stb.ch/announce/cc or send an email to licensetstb-stb.ch).
CC	
DR	EMBL: X54131; CMA38066.1; -
DR	PfR; SI2050; SI2050.
DR	HSSP; P18052; 1yfo.
DR	MIM; 176882; -
DR	InterPro: IPRO03961; FN_III.
DR	InterPro: IPRO00387; TYR_phosphatase.
DR	InterPro: IPRO00242; Tyr_prot_Phpntase.
DR	Pfam: PF00041; fn3; 16.
DR	Pfam: PF00102; Y-phosphatase; 1.
DR	PRINTS; PR00700; PTYPHPPTASE.
DR	SMART; SMART0060; FN3; 15.
DR	SMART; SMART0194; PTPc; 1.
DR	PROSITE; PS00383; TYR_PHOSPATASE_1; 1.
DR	PROSITE; PS00056; TYR_PHOSPATASE_2; 1.
DR	PROSITE; PS00053; TYR_PHOSPATASE_PPp; 1.
KW	Glycoprotein; Transmembrane; Hydrolase; Phosphorylation; Repeat; Signal.
KW	SIGNAL.
FT	CHAIN 1..22
FT	DOMAIN 23..1997
FT	DOMAIN 23..1621
FT	TRANSMEM 1622..1642
FT	DOMAIN 1643..1997
FT	DOMAIN 23..110
FT	DOMAIN 111..200
FT	DOMAIN 201..286
FT	DOMAIN 287..374
FT	DOMAIN 375..464
FT	DOMAIN 465..551
FT	DOMAIN 552..640
FT	DOMAIN 641..728
FT	DOMAIN 729..816
FT	DOMAIN 817..904
FT	DOMAIN 905..992
FT	DOMAIN 993..1082
FT	DOMAIN 1083..1170
FT	DOMAIN 1171..1268
FT	DOMAIN 1269..1352
FT	DOMAIN 1353..1442
FT	DOMAIN 1443..1529
FT	DOMAIN 1530..1616
FT	ACT_SITE 1904..1904
FT	CARBOHYD 28..28
FT	CARBOHYD 53..53
FT	CARBOHYD 75..75
FT	CARBOHYD 172..172
FT	CARBOHYD 198..198
FT	CARBOHYD 267..267
FT	CARBOHYD 321..321
FT	CARBOHYD 414..414
FT	CARBOHYD 421..421
FT	CARBOHYD 479..479
FT	CARBOHYD 544..544
FT	CARBOHYD 544..544

FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 598 598 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 652 652 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 829 829 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1096 1096 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1163 1163 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1185 1185 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1212 1212 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1274 1274 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1367 1367 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1470 1470 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1474 1474 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1518 1518 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 1997 AA: 224267 MW: 691E99BA7A1515DD CRC64;

Query Match 52.8%; Score 38; DB 1; Length 1997;
 Best Local Similarity 50.0%; Pred. No. 99;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GEIQTQPDV 10
 Db 1349 GSVRTKPKDI 1358

RESULT 5
 MCH_HALNT1 STANDARD; PRT; 311 AA.
 ID MCH_HALNT1
 AC 09HPD7;
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT N5,N10-methenyltetrahydromethanopterin cyclohydrolase (EC 3.5.4.27)
 DE (Methenyl-H4MPT cyclohydrolase).
 GN MCH OR VNG1686G.
 OS Halobacterium sp. (strain NRC-1).
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 OC Halobacterium.
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20504483; PubMed=11016950;
 RA NG W.V., Kennedy S.P., Mahaitas G.G., Bergquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sirogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonksi P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Edhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
 RT "Genome sequence of Halobacterium species NRC-1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 CC -1- FUNCTION: Reversible interconversion of N(5)-formyl-H(4)MPT to
 CC methenyl-H(4)MPT(+) (by similarity).
 CC -1- CATALYTIC ACTIVITY: 5,10-methenyl-5,6,7,8-tetrahydromethanopterin
 CC + H(2)O = N5-formyl-5,6,7,8-tetrahydromethanopterin.
 CC -1- PATHWAY: Methanogenesis (reduction of carbon dioxide to methane);
 CC third step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE MCH FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE005076; AAG19933.1; -
 CC DR HSSP: P94954; 10LM.

DR InterPro: IPR003209; MCH.
 DR Pfam: PF02289; MCH; 1.
 KM Hydrolase; Methanogenesis; Complete proteome.
 SO SEQUENCE 311 AA: 32140 MW: 57A250AAE0B2D3AE CRC64;

Query Match 51.4%; Score 37; DB 1; Length 311;
 Best Local Similarity 58.3%; Pred. No. 21;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Oy 3 IOTKPDVSGAT 14
 Db 63 VQTRVDTVGDPAT 74

RESULT 6
 ZIC1_XENLA STANDARD; PRT; 443 AA.
 ID ZIC1_XENLA
 AC 073689; 057461;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Zinc finger protein ZIC 1 (Zinc finger protein of the cerebellum 1)
 DE (ZIC-related 1 protein) (ZIC-R1) (ODD-paired-like).
 GN ZIC1 OR ZIC-1 OR OPL.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=9355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98167885; PubMed=9435279;
 RA Mizusaki K., Kishi M., Matsui M., Nakanishi S., Sasai Y.;
 RT "Xenopus Zic-related-1 and Sox-2, two factors induced by chordin, have
 RT distinct activities in the initiation of neural induction.";
 RL Development 125:579-587(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98322239; PubMed=9655809;
 RA Kuo J.S., Patel M., Gamse J., Merzdorf C., Liu X., Apekin V., Sive H.;
 RT "Opl: a zinc finger protein that regulates neural determination and
 RT patterning in Xenopus.";
 RL Development 125:2867-2882(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98413039; PubMed=9739105;
 RA Nakata K., Nagai T., Aruga J., Mikoshiba K.;
 RT "Xenopus Zic family and its role in neural and neural crest
 RT development.";
 RL Mech. Dev. 75:43-51(1998).
 CC -1- FUNCTION: MAY PLAY A ROLE IN CEREBELLAR DEVELOPMENT. CAN ACTIVATE
 CC THE PROMOTORAL GENE NEUROGENIN-RELATED-1, AND INITIATE NEURAL AND
 CC NEURONAL DIFFERENTIATION. MAY MEDIANE NEURAL COMPETENCE AND MAY BE
 CC INVOLVED IN ACTIVATION OF MIDBRAIN, DORSAL NEURAL AND NEURAL CREST
 CC FATES. CAN SENSITIZE THE PRESUMPTIVE NEUROECTODERM FOR INDUCTION,
 CC SUGGESTING THAT IT MAY BE A NEURAL COMPETENCE FACTOR. AN ACTIVATED
 CC FORM OF OPL CAN ACTIVATE NEURAL CREST AND DORSAL NEURAL TUBE
 CC MARKERS. CAN SYNERGIZE WITH NOGIN TO INDUCE EXPRESSION OF
 CC ENGRAILED.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: CNS. A HIGH LEVEL EXPRESSION IS SEEN IN THE
 CC CEREBELLUM.
 CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED AT THE BEGINNING OF
 CC GASTRULATION. IT IS ACTIVATED ZYGOTICALLY BY LATE BLASTULA. ITS
 CC EXPRESSION IS MAXIMAL BY MIDGASTRULA, WITH LOWER EXPRESSION
 CC PERSISTING INTO TAILBD STAGES. DURING EARLY GASTRULA STAGES IT IS
 CC EXPRESSED WIDELY IN THE DORSAL ECTODERM. WHEN NEURAL INDUCTION
 CC OCCURS, IT HAS WIDE AND HOMOGENOUS EXPRESSION IN THE PROSPECTIVE
 CC NEUROECTODERM. LATER, AT LATE GASTRULA STAGES, ITS EXPRESSION IS
 CC SHUT OFF IN THE DORSAL MIDLINE EXCEPT IN THE MOST ANTERIOR AREA.
 CC AT NEURAL PLATE STAGES, IT IS EXPRESSED IN THE LATERAL PART OF
 CC ANTERIOR NEURAL PLATE AND IN THE FLANKING ECTODERM. AT LARVAL

```
CC STAGES, IT IS DETECTED MAINLY IN THE DORSAL NEURAL TUBE THROUGHOUT
CC ITS A-P AXIS.
CC -1- DOMAIN: HAS A CARBOXY TERMINAL REGULATORY DOMAIN, WHICH WHEN
CC REMOVED INCREASES OPL ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF022927; AAC14214.1; -
CC DR EMBL: AF028805; AAB99946.1; -
CC DR EMBL: AB009564; BAA33406.1; -
CC DR HSSP: P08047; 1SR2.
CC DR InterPro: IPR000822; Znf-C2H2.
CC DR Pfam: PF00096; ZF-C2H2; 5.
CC DR PRINTS: PR00048; ZINC_FINGER.
CC DR SMART: SM00355; Znf_C2H2; 4.
CC DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
CC DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 4.
CC DR Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
CC Transcription regulation; Activator; Developmental protein;
CC Neurogenesis.
CC KW DOMAIN 71 74 POLY-ALA.
CC FT DOMAIN 77 80 POLY-HIS.
CC FT ZN_FING 221 380 ZINC_FINGERS.
CC FT ZN_FING 221 380 C2H2-TYPE (ATYPICAL).
CC FT ZN_FING 265 292 C2H2-TYPE (ATYPICAL).
CC FT ZN_FING 298 322 C2H2-TYPE.
CC FT ZN_FING 328 352 C2H2-TYPE.
CC FT ZN_FING 358 380 C2H2-TYPE.
CC FT DOMAIN 408 443 INCREASES ACTIVITY WHEN MISSING.
CC FT CONFLICT 5 5 A -> T (IN REF. 2).
CC FT SEQUENCE 443 AA; 48252 MW; E61CA481CFAC106 CRC64;

Query Match 51.4%; Score 37; DB 1; Length 443;
Best Local Similarity 42.9%; Pred. NO. 30;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 GEIQKPRVGOAT 14
ID RRPV_CVMD STANDARD; PRT; 500 AA.
AC P26627;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE RNA-directed RNA polymerase (ORF1A) (EC 2.7.7.48) (Fragment).
OS Murine coronavirus MHV (strain defective JHM).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OC NCBI_TaxID=11143;
RX MEDLINE=91056574; Pubmed=2243386;
RA Makino S., Yokomori K., Lai M.M.C.;
RT "Analysis of efficiently packaged defective interfering RNAs of
RT murine coronavirus: localization of a possible RNA-packaging
RT signal."
RL J. Virol. 64:6045-6053(1990).
CC -1- FUNCTION: THE RNA DEPENDENT RNA POLYMERASE OF CORONAVIRUSES IS
CC A MULTIFUNCTIONAL PROTEIN: IT CONTAINS THE ACTIVITIES NECESSARY
CC FOR THE TRANSCRIPTION OF NEGATIVE STRANDED RNA, LEADER RNA,
```

```
CC SUBGENOMIC MRNAs AND PROGENY VIRION RNA.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC (RNA)(N).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M61144; AAA46451.1; -
CC DR PIR: A36388; A36388.
CC DR HSSP: P13257; 1LF8.
CC DR InterPro: IPR002705; Peptidase_C16.
CC DR Pfam: PF01831; Peptidase_C16; 1.
CC KW Transferase; RNA-directed RNA polymerase.
CC FT NON_TER 500 500
CC FT SEQUENCE 500 AA; 55246 MW; B24DB851ED951264 CRC64;

Query Match 51.4%; Score 37; DB 1; Length 500;
Best Local Similarity 50.0%; Pred. NO. 35;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 GEIQKPRVGOAT 14
ID R18E_HALMA STANDARD; PRT; 115 AA.
AC P12733;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L18E (HL29) (L19).
DE RPL18E.
OS Haloarcula marismortui (Haloacterium marismortui).
OC Archaea; Euryarchaeota; Haloacteriales; Haloacteriaceae; Haloarcula.
OC NCBI_TaxID=2238;
OX [1]
RX MEDLINE=92105119; Pubmed=1840597;
RA Kroeber W.J., Arndt E.;
RT "Haloacetal S9 operon. Three ribosomal protein genes are
RT cotranscribed with genes encoding a tRNA(Leu), the enolase, and a
RT putative membrane protein in the archaeobacterium Haloarcula
RT (Haloacterium) marismortui."
RL J. Biol. Chem. 266:24573-24579(1991).
RN [2]
RN SEQUENCE:
RX MEDLINE=8816754; Pubmed=3350019;
RA Hatakeyama T., Kimura M.;
RT "Complete amino acid sequences of the ribosomal proteins L25, L29 and
RT L31 from the archaeobacterium Haloacterium marismortui."
RL Eur. J. Biochem. 172:703-711(1988).
RN [3]
RN SEQUENCE OF 1-71 FROM N.A.
RX MEDLINE=92291093; Pubmed=1376318;
RA Scholzen T., Arndt E.;
RT "The alpha operon equivalent genome region in the extreme halophilic
RT archaeobacterium Haloarcula (Haloacterium) marismortui."
RL J. Biol. Chem. 267:12123-12130(1992).
RN [4]
RN SEQUENCE OF 1-23.
RX MEDLINE=89062418; Pubmed=3196689;
RA Walsh M.J., McDougall J., Wilmann-Liebold B.;
RT "Extended N-terminal sequencing of proteins of archaeobacterial
RT ribosomes plotted from two-dimensional gels onto glass fiber and
RT poly(vinylidene difluoride) membrane.";
```

```

RL Biochemistry 27:6867-6876(1988).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RC STRAIN-ATCC 43049;
RX MEDLINE-20396344; PubMed-10937989;
RA Ban N., Nissen P., Hansen J., Moore P.B., Steltz T.A.;
RT "The complete atomic structure of the large ribosomal subunit at 2.4
  A resolution."
RL Science 289:905-920(2000).
CC -1- SIMILARITY: BELONGS TO THE L18E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
CC EMBL: M76567; AAA73096.1; -
CC EMBL: M87833; AAA73213.1; -
CC PIR: A41715; RSHSH9.
CC PDB: 1FEK; 14-AUG-00.
CC InterPro: IPR000039; Ribosomal_L18e.
CC Pfam: PF00256; L15; 1
CC PROSITE: PS01106; RIBOSOMAL_L18E; 1.
CC Ribosomal protein; 3D-structure.
CC INIT_MER 0
CC SEQUENCE 115 AA; 12291 MW; 1FEID7368A7E2AFB CRC64;

Query Match          50.0%; Score 36; DB 1; Length 115;
Best Local Similarity 53.8%; Pred. No. 11;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 GEIOTKPDVGOA 13
   ||| ||| |||
Db 83 GTAFETKIDYGEA 95

RESULT 9
DDBD_ASPNG          STANDARD;          PRT;          292 AA.
ID DBD ASPNG
AC P80346;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE O-pyrocatechuate decarboxylase (EC 4.1.1.46) (2,3-dihydroxybenzoic
DE acid decarboxylase) (DDBD) (Fragments)..
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiatales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE.
RX MEDLINE-95324511; PubMed-7601088;
RA Senthia R., Savithri H.S., Rao N.A., Vaidyanathan C.S.;
RT "2,3-dihydroxybenzoic acid decarboxylase from Aspergillus niger. A
RT novel decarboxylase."
RL Eur. J. Biochem. 230:104-110(1995).
CC -1- CATALYTIC ACTIVITY: 2,3-dihydroxybenzoate + catechol + CO(2).
CC -1- PATHWAY: LAST STEP IN THE METABOLISM OF INDOLE TO CATECHOL.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- MISCELLANEOUS: THE MW OF THE COMPLETE PROTEIN IS ABOUT 38 kDa (345
CC RESIDUES).
CC Lyase; Decarboxylase.
KM ACT_SITE 263
FT ACT_SITE 263
FT UNSURE 247 OR G.
FT UNSURE 250 OR G.
FT NON_CONS 32
FT NON_CONS 33
FT NON_CONS 66
FT NON_CONS 88
FT NON_CONS 121
FT NON_CONS 122

```

```

FT NON_CONS 150
FT NON_CONS 160
FT NON_CONS 161
FT NON_CONS 170
FT NON_CONS 171
FT NON_CONS 177
FT NON_CONS 183
FT NON_CONS 184
FT NON_CONS 194
FT NON_CONS 195
FT NON_CONS 207
FT NON_CONS 208
FT NON_CONS 214
FT NON_CONS 215
FT NON_CONS 224
FT NON_CONS 225
FT NON_CONS 229
FT NON_CONS 230
FT NON_CONS 236
FT NON_CONS 237
FT NON_CONS 240
FT NON_CONS 241
FT NON_CONS 247
FT NON_CONS 248
FT NON_CONS 255
FT NON_CONS 256
FT NON_CONS 264
FT NON_CONS 265
FT NON_CONS 292
FT NON_CONS 292
SQ SEQUENCE 292 AA; 33000 MW; EF78074854BAD925 CRC64;

Query Match          50.0%; Score 36; DB 1; Length 292;
Best Local Similarity 53.8%; Pred. No. 30;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EIOTKPDVGOAT 14
   ||| ||| |||
Db 129 QIRDKPDVFGAFT 141

RESULT 10
Y825_STRCO          STANDARD;          PRT;          340 AA.
ID Y825_STRCO
AC 005954; O9KZU3;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical transcriptional regulator SCD84.25C.
GN SCD84.25C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Saunders D.C., Harris D., Cerdano A.M., Parkhill J., Barrell B.G.,
RA Rajandream M.A.; (Apr-2000) to the EMBL/GenBank/DBJ databases.
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-66 FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE-93345814; PubMed-7688332;
RA Wray L.V. Jr., Fisher S.H.;
RT "The Streptomyces coelicolor glrK gene encodes a protein similar to
RT other bacterial response regulators."
RL Gene 130:145-150(1993).
CC -1- SIMILARITY: BELONGS TO THE LACI FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
CC EMBL: AL353616; CAB88491.1; -
CC EMBL: L03213; AAA02839.1; -
CC PIR: P06444; PNO644.
CC HSSP: P15039; 1PRU.
CC InterPro: IPR000843; HTH_Laci.
CC Pfam: PF00532; Peripla_BP_Like; 1.
CC SMART: SM00354; HTH_LACI; 1.
CC PROSITE: PS00356; HTH_LACI_FAMILY; FALSE_NEG.

```

KW Hypothetical protein: Transcription regulation; DNA-binding.
 FT DNA BIND 25 H-T MOTIF (POTENTIAL).
 SQ SEQUENCE 340 AA: 36766 MW: 35F32456E94E1A15 CRC64;
 Query Match 50.0%; Score 36; DB 1; Length 340;
 Best Local Similarity 58.3%; Pred. NO. 35;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 2 EIQTKPDVVGQA 13
 I: ||||| I
 Db 45 ELGYRDPVQA 56
 RESULT 11
 HYP_HUMAN STANDARD; PRT; 455 AA.
 ID HYP_HUMAN STANDARD; PRT; 455 AA.
 AC P07099;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Epoxide hydrolase 1 (EC 3.3.2.3) (Microsomal epoxide hydrolase)
 DE (Epoxide hydrolase).
 GN EPHX1 OR EPHX OR EPOX.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCB1_TaxID=9606;
 RX MEDLINE-8801564; AND SEQUENCE OF 1-19.
 RX MEDLINE-88087301; PubMed-2891713;
 RX Skoda R.C., Demierre A., McFried O.W., Gonzalez F.J., Meyer U.A.;
 "Human microsomal xenobiotic epoxide hydrolase. Complementary DNA
 sequence, complementary DNA-directed expression in COS-1 cells, and
 chromosomal localization.";
 J. Biol. Chem. 263:1549-1554(1988).
 [2]
 SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RL Wilson N.M., Omiecinski C.J.;
 RL Submitted (JUL-1988) to the EMBL/Genbank/DBJ databases.
 [3]
 SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE-8801564; PubMed-3502697;
 RX Jackson M.R., Craft J.A., Burchell B.;
 "Nucleotide and deduced amino acid sequence of human liver microsomal
 epoxide hydrolase.";
 Nucleic Acids Res. 15:7188-7188(1987).
 [4]
 SEQUENCE FROM N.A., AND VARIANTS HIS-113; ARG-139 AND ILE-396.
 RC TISSUE=Liver;
 RX MEDLINE-9428203; PubMed-7516776;
 RX Hassett C., Aicher L., Sidhu J.S., Omiecinski C.J.;
 "Human microsomal epoxide hydrolase: genetic polymorphism and
 functional expression in vitro of amino acid variants.";
 Hum. Mol. Genet. 3:421-428(1994).
 [5]
 ERRATUM.
 RA Hassett C., Aicher L., Sidhu J.S., Omiecinski C.J.;
 Hum. Mol. Genet. 3:1214-1214(1994).
 [6]
 SEQUENCE FROM N.A.
 RX MEDLINE-9513759; PubMed-7835893;
 RA Hassett C., Robinson K.B., Beck N.B., Omiecinski C.J.;
 "The human microsomal epoxide hydrolase gene (EPHX1): complete
 nucleotide sequence and structural characterization.";
 Genomics 23:433-442(1994).
 [7]
 SEQUENCE OF 9-327 FROM N.A.
 RP TISSUE=Liver;
 RC Craft J.A., Jackson M.R., Burchell B.;
 "Partial nucleotide sequence of a cloned cDNA for human liver

RT microsomal epoxide hydrolase.";
 RL Biochem. Soc. Trans. 15:708-709(1987).
 CC -1- FUNCTION: BIOTRANSFORMATION ENZYME THAT CATALYZES THE HYDROLYSIS
 CC OF ARENE AND ALIPHATIC EPOXIDES TO LESS REACTIVE AND MORE WATER
 CC SOLUBLE DIHYDRODIOLS BY THE TRANS ADDITION OF WATER.
 CC -1- CATALYTIC ACTIVITY: An epoxide + H(2)O = a glycol.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE BOUND MICROSOMES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S33.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J03518; AAA61305.1; -
 DR EMBL: X07936; CAA30759.1; -
 DR EMBL: Y00424; CAA68486.1; -
 DR EMBL: M36374; AAA59580.1; -
 DR EMBL: L25878; AAA52389.1; -
 DR EMBL: L25879; AAA52390.1; -
 DR EMBL: AF253417; AAC41694.1; -
 DR EMBL: U06661; AAB60649.1; -
 DR EMBL: U06657; AAB60649.1; JOINED.
 DR EMBL: U06657; AAB60649.1; JOINED.
 DR EMBL: U06658; AAB60649.1; JOINED.
 DR EMBL: U06659; AAB60649.1; JOINED.
 DR EMBL: U06660; AAB60649.1; JOINED.
 DR PIR: S03114; S03114.
 DR PIR: A26856; A26856.
 DR PIR: A29939; A29939.
 DR MEROPS: S33.971; -
 DR MIM: 132810; -
 DR InterPro: IPR000073; Abhydrolase.
 DR InterPro: IPR000639; Epox hydrolase.
 DR InterPro: IPR000379; Est_lip_thioest_actsite.
 DR Pfam: PF00561; abhydrolase_1.
 DR PRINTS: PR00412; EPOXHYDRLASE.
 KW Hydrolase; Endoplasmic reticulum; Detoxification; Transmembrane;
 KW Liver; Aromatic hydrocarbons catabolism; Microsome; Polymorphism.
 FT TRANSMEM 2 20
 FT SIGNAL-ANCHOR (POTENTIAL).
 FT Y -> H (FREQUENT ALLELE).
 FT /FTID-VAR_005295.
 FT VARIANT 139 139
 FT H -> R (FREQUENT ALLELE).
 FT /FTID-VAR_005296.
 FT T -> I (EITHER AN INFREQUENT POLYMORPHISM
 FT OR A SEQUENCING ERROR).
 FT /FTID-VAR_005297.
 FT Y -> H: PARTIAL LOSS OF ENZYME ACTIVITY.
 FT H -> R: SMALL INCREASE IN ENZYME ACTIVITY.
 FT R -> K (IN REF. 3).
 FT H -> N (IN REF. 3 AND 7).
 FT V -> L (IN REF. 3).
 FT K -> S (IN REF. 3).
 FT L -> F (IN REF. 3).
 FT L -> V (IN REF. 3).
 FT CONFLICT 406 406
 FT CONFLICT 420 420
 FT L -> V (IN REF. 3).
 FT SEQUENCE 455 AA: 52949 MW: 88E333838C841390 CRC64;
 QY 5 TKPDVVGQA 13
 I: ||||| I
 Db 305 TKPDVVGQA 313
 RESULT 12
 AF6_HUMAN STANDARD; PRT; 1816 AA.
 ID AF6_HUMAN STANDARD; PRT; 1816 AA.
 Query Match 50.0%; Score 36; DB 1; Length 455;
 Best Local Similarity 77.8%; Pred. NO. 48;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

AC P55196; 075087; 075088; 075089; Q9NU92;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE AF-6 protein.
 GN MLLT4 OR AF6.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=94061833; PubMed=8242616;
 RA Prasad R., Gu Y., Alder H., Nakamura T., Canaan O., Saito H.,
 Huebner K., Gale R.P., Nowell P.C., Kuriyama K., Miyazaki Y.,
 Croce C.M., Canaan E.;
 RT "Cloning of the ALF-1 fusion partner, the AF-6 gene, involved in
 RT acute myeloid leukemias with the t(6;11) chromosome translocation."
 RL Cancer Res. 53:5624-5628(1993).
 RN [2]
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
 RX TISSUE=Fetal brain;
 RX MEDLINE=98344142; PubMed=9679199;
 RA Saito S., Matsushima M., Shirahama S., Minaguchi T., Kanamori Y.,
 Minami M., Nakamura Y.;
 RT "Complete genomic structure, DNA polymorphisms, and alternative
 RT splicing of the human AF-6 gene."
 RL DNA Res. 5:115-120(1998).
 RN [3]
 RP SEQUENCE OF 337-1816 FROM N.A. (ISOFORM 2).
 RA Williams S.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY ACT AS AN INTRACELLULAR SIGNALING COMPONENT
 CC -1- CONTROLLED BY RAS SIGNALING PATHWAYS.
 CC -1- SOBUNIT: BINDS DIRECTLY TO ZO-1 AND OCCUDIN.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (probable).
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1, 2 (SHOWN HERE) AND 3; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY A CHROMOSOMAL
 CC TRANSLOCATION T(6;11)(Q27:Q23) THAT INVOLVES MLLT4 AND MLL/HRX.
 CC THE RESULT IS A ROGUE ACTIVATOR PROTEIN.
 CC -1- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 FHA DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB011399; BAA32484.1; -
 DR EMBL; AB011399; BAA32483.1; -
 DR EMBL; AB011399; BAA32485.1; -
 DR EMBL; U02478; AAC50059.1; -
 DR EMBL; AL049698; CAB76850.1; -
 DR HSP; Q12923; 3PDZ.
 DR MIM; 159559; -
 DR InterPro: IPR002710; DIL.
 DR InterPro: IPR000253; FHA_domain.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR000159; RA.
 DR Pfam; PF01843; DIL; 1.
 DR Pfam; PF00498; FHA; 1.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00788; RA; 2.
 DR ProDom; PD003376; DIL; 1.
 DR SMART; SM00240; FHA; 1.
 DR SMART; SM00228; PDZ; 1.
 DR SMART; SM00314; RA; 2.
 DR PROSITE; PSS0106; PDZ; 1.

KW Chromosomal translocation; Proto-oncogene; Alternative splicing.
 FT DOMAIN 36 206
 FT DOMAIN 425 491
 FT DOMAIN 804 910
 FT DOMAIN 991 1077
 FT DOMAIN 162 174
 FT DOMAIN 1349 1356
 FT DOMAIN 1371 1376
 FT DOMAIN 1561 1571
 FT SITE 26
 FT
 FT VARSPLIC 1588 1611
 FT VARSPLIC 1612 1816
 FT VARSPLIC 1666 1743
 FT
 FT VARSPLIC 1744 1816
 FT CONFLICT 373 373
 FT CONFLICT 391 391
 FT CONFLICT 744 744
 FT CONFLICT 1031 1031
 FT CONFLICT 1408 1408
 SQ SEQUENCE 1816 AA; 205604 MW; EB1FE7F04879CEBF CRC64;
 Query Match 50.0%; Score 36; DB 1; Length 1816;
 Best Local Similarity 75.0%; Pred. No. 2,1e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 EIOTKPDR 9
 Db 1525 ELOSKPDR 1532
 ID NM1_YEAST STANDARD; PRT; 2748 AA.
 AC 000402;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Nuclear migration protein NM1.
 GN NM1 OR YDR150M.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SRRAIN-ATCC 28383 / FL100;
 RX MEDLINE=92079907; PubMed=1745235;
 RA Kornmeier J., Schaeff-Gerstenschlaeger I., Zimmermann F.K.,
 RA Perleke D., Kuentzel H.;
 RT "Nuclear migration in Saccharomyces cerevisiae is controlled by the
 RT highly repetitive 313 kDa NM1 protein."
 RL Mol. Gen. Genet. 230:277-287(1991).
 CC -1- FUNCTION: CONTROLS NUCLEAR MIGRATION. NM1 SPECIFICALLY CONTROLS
 CC THE INTERACTION OF THE BUD NECK CYTOSKELETON WITH THE PRE-
 CC DIVISIONAL G2 NUCLEUS PERHAPS BY RECOGNIZING G2-SPECIFIC
 CC CYTOPLASMIC MICROTUBULI OR OTHER COMPONENTS OF THE NUCLEAR
 CC ENVELOPE.
 CC -1- MISCELLANEOUS: ADDITIONAL REGIONS OF LOWER HOMOLOGY TO THE REPEAT
 CC CONSENSUS (ALWAYS STARTING WITH PROLINE) ARE FOUND IN BOTH
 CC FLANKING DOMAINS OF THE TANDDEM REPEATS.
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 IS THE INITIATOR.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: X61236; CAA43554.1; -
DR PIR: S19052; S19052.
DR SCD: S000257; NDM1.
DR InterPro: IPR001849; PH.
DR Pfam: PF00169; PH: 1.
DR SMART: SM00233; PH: 1.
DR PROSITE: PS50003; PH: 1.
DR Repeat: -
FT DOMAIN 1.
FT REPEAT 593 1384
FT REPEAT 593 656
FT REPEAT 657 727
FT REPEAT 728 798
FT REPEAT 799 862
FT REPEAT 863 926
FT REPEAT 927 990
FT REPEAT 991 1054
FT REPEAT 1055 1118
FT REPEAT 1119 1182
FT REPEAT 1183 1246
FT REPEAT 1247 1310
FT REPEAT 1311 1374
FT REPEAT 1375 1384
FT DOMAIN 2573 2683
SO SEQUENCE 2748 AA; 313202 MW; B2FBD67C9F6211AE CRC64;

Query Match
Best Local Similarity 50.08; Score 36; DB 1; Length 2748;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 IOTKPRVGOAT 14
Db 2441 LTKEDKKGQAT 2452

RESULT 14
ID THB_PAROL STANDARD; PRT; 395 AA.
AC Q91279;
DT 15-JUL-1999 (Rel. 38. Created)
DT 15-JUL-1999 (Rel. 38. Last sequence update)
DE Thyroid hormone receptor beta (THR-beta).
GN THRB OR NR1A2.
OS Paracanthus olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Pleuronectidae; Acanthopterygii; Percomorpha; Pleuronectiformes; NCB1_TaxID=8255;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-96149097; PubMed-8536930;
RA Yamano K., Inui Y.;
RT "CDNA cloning of thyroid hormone receptor beta for the Japanese flounder";
RL Gen. Comp. Endocrinol. 99:197-203(1995).
CC -1- FUNCTION: HIGH AFFINITY RECEPTOR FOR TRIIODOTHYRONINE.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; BETA-1 (SHOWN HERE) AND BETA-2;
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN, A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: D45245; BAA08201.1; -
DR HSSP: P10828; 2NL.
DR InterPro: IPR000536; Hormone_rec_1lg.
DR InterPro: IPR001723; Sterdormone_receptor.
DR Pfam: PF00104; hormone_rec. 1.
DR PRINTS: PR00398; STROHOMER.
DR PRINTS: PR00447; STROIDFINGER.
DR SMART: SM00430; ZNF_C4; 1.
DR PROSITE: PS00399; ZNF_C4; 1.
DR Receptor: Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger; Multigene family; Alternative splicing.
KW DOMAIN 1 31
FT DNA_BIND 32 99
FT ZN_FING 32 52
FT ZN_FING 70 94
FT DOMAIN 149 395
SO SEQUENCE 395 AA; 45113 MW; F4319CD96FE2451 CRC64;

Query Match
Best Local Similarity 48.6%; Score 35; DB 1; Length 395;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 GIKTKPRVGOAT 14
Db 175 GVKEDKPEIGAS 188

RESULT 15
ID HYDH_ECOLI STANDARD; PRT; 465 AA.
AC P14377;
DT 01-JAN-1990 (Rel. 13. Created)
DT 01-OCT-1993 (Rel. 27. Last sequence update)
DE Sensor protein hydH (Ec 2.7.3.-).
GN HYDH OR B4003.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
OX NCB1_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RA Blatter F.R., Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.;
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the Nucleic Acids Res. 21:5408-5417(1993).
RN [2]
RP SEQUENCE OF 328-465 FROM N.A.
RC STRAIN-K12;
RX MEDLINE-89327164; PubMed-2666400;
RA Stoker K., Reijnders W.N.M., Oltmann L.F., Stoutamer A.H.;
RT "Initial cloning and sequencing of hydH, an operon homologous to ntrBC and regulating the labile hydrogenase activity in Escherichia coli K-12";
RL J. Bacteriol. 171:4448-4456(1989).
CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM HYDH/HYDG INVOLVED IN THE REGULATION OF THE LABILE HYDROGENASE ACTIVITY.
CC HIGH MAY FUNCTION AS A MEMBRANE-ASSOCIATED PROTEIN KINASE THAT PHOSPHORYLATES HYDG IN RESPONSE TO ENVIRONMENTAL SIGNALS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner memb-----

```

CC (Probable).
CC -1 SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U00006; AAC43101.1;
DR EMBL: AE000473; AAC76977.1;
DR EMBL: M28369; AAA24003.1;
DR PIR: A33862; A33862.
DR Ecogene: EG10008; byGH.
DR InterPro: IPR004358; BCTRLSENSOR.
DR InterPro: IPR003594; HATPase_C.
DR InterPro: IPR004359; HIS_KIN_s19.
DR InterPro: IPR003661; HIS_KIN_A.
DR Pfam: PF02518; HATPase_C_1.
DR Pfam: PF00512; signal_1.
DR PRINTS: PR00344; BCTRLSENSOR.
DR SMART: SM00387; HATPase_C_1.
DR SMART: SM00388; HIS_KIN_1.
DR PROSITE: PS50109; HIS_KIN_1.
DR Sensory transduction: Transferase; Kinase; Phosphorylation;
KW Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 15 35 POTENTIAL.
FT TRANSMEM 84 104 POTENTIAL.
FT TRANSMEM 202 222 POTENTIAL.
FT TRANSMEM 360 380 POTENTIAL.
FT TRANSMEM 407 427 POTENTIAL.
FT TRANSMEM 458 458 POTENTIAL.
FT DOMAIN 251 458 HISTIDINE KINASE.
FT CONFLICT 361 361 L->V (IN REF. 2).
FT CONFLICT 383 387 SESGA -> TRAG (IN REF. 2).
SQ SEQUENCE 465 AA: 51031 MW: 0472AC3494E95EE3 CRC64;

```

```

Query Match 48.6%; Score 35; DB 1; Length 465;
Best Local Similarity 63.6%; Pred. NO. 75;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
OY 2 EIQTKPDRVQ 12
DB 349 EIQADPDRITQ 359

```

Search completed: June 10, 2002, 15:29:41
Job time: 476 sec

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 PS
 XX Claim 13; SEQ ID 4314; 71bp + CD-ROM; English.
 CC The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.
 CC
 SQ Sequence 100 AA:

Query Match 47.5%; Score 38; DB 21; Length 100;
 Best Local Similarity 77.8%; Pred. No. 47;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 SDGFTYTMEN 10
 : || ||||:
 Db 61 sdgytymes 69

RESULT 15
 AAG90736
 ID AAG90736 standard; Protein: 135 AA.
 AC AAG90736;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum protein fragment SEQ ID NO: 4490.
 XX
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX
 OS Corynebacterium glutamicum.
 OS
 XX
 PN EPI108790-A2.
 PN
 XX
 PD 20-JUN-2001.
 PD
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 PF
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 PR
 XX
 PA (RYOW) KYOWA HAKKO KOGYO KK.
 PA
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI: 2001-376931/40.
 DR N-PSDB: AAH65955.
 DR
 XX
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 PS Claim 17; SEQ ID NO: 4490; 246bp + Sequence Listing; English.
 PS
 XX

CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 CC
 SQ Sequence 135 AA:

Query Match 47.5%; Score 38; DB 22; Length 135;
 Best Local Similarity 60.0%; Pred. No. 65;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MSDGFTYTMEN 10
 : ||||:
 Db 94 lssgfttven 103

Search completed: June 10, 2002, 15:21:08
 Job time: 168 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 15:22:24 ; Search time 35.61 Seconds
(without alignments)
40.476 Million cell updates/sec

Title: US-09-647-522-3

Perfect score: 80

Sequence: 1 MSDGFTTMSDRRK 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:**
2: p1r2:**
3: p1r3:**
4: p1r4:**

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	100.0	450	2 JC7371	toxin - jellyfish
2	43	53.8	550	2 T10393	chitinase (EC 3.2.
3	42	52.5	552	2 S36786	carboxylesterase (
4	41	51.2	214	2 D90185	SSU ribosomal prot
5	41	51.2	387	2 D70090	hypothetical prote
6	41	51.2	2550	2 B53435	vesicular transport
7	40	50.0	376	2 C97367	recf protein (lipo
8	40	50.0	376	2 AD2585	DNA replication an
9	40	50.0	437	2 C75632	probable hemolysin
10	40	50.0	704	2 H82012	capsule polysaccha
11	40	50.0	704	2 D81240	capsule polysaccha
12	40	50.0	833	2 J01866	hypothetical 87.1k
13	39	48.8	160	2 G95213	hypothetical prote
14	39	48.8	164	2 H98077	hypothetical prote
15	39	48.8	187	2 E84263	glycerol-3-phospha
16	39	48.8	339	2 T07853	probable fructose-
17	39	48.8	357	2 S63142	hypothetical prote
18	39	48.8	360	2 H97703	recf protein (lipo
19	39	48.8	387	2 T48621	hypothetical prote
20	39	48.8	411	2 S58094	hypothetical KM do
21	39	48.8	438	2 E84579	hypothetical prote
22	39	48.8	572	1 DEECPC	pyruvate dehydroge
23	39	48.8	572	2 E90748	pyruvate oxidase I
24	39	48.8	572	2 A85599	pyruvate oxidase P
25	39	48.8	572	2 AG0608	pyruvate dehydroge
26	39	48.8	611	2 G96032	probable thiamine
27	38	47.5	138	2 H83299	hypothetical prote
28	38	47.5	354	1 S17953	alkanal monooxygen
29	38	47.5	360	2 E71710	recf protein (recf

30	38	47.5	464	2 T38699	glutathione reduct
31	38	47.5	520	2 T06625	hypothetical prote
32	38	47.5	687	2 G81970	probable glycine--
33	38	47.5	687	2 E81027	glycyl-tRNA synth
34	37	46.2	49	2 D97048	hypothetical prote
35	37	46.2	210	1 Z78PT9	gene 55.10 protein
36	37	46.2	221	2 H84781	hypothetical prote
37	37	46.2	228	2 T51147	hypothetical prote
38	37	46.2	265	2 D82871	conserved hypothet
39	37	46.2	421	2 A12090	two-component resp
40	37	46.2	520	2 S14599	E2 glycoprotein -
41	37	46.2	520	2 S14600	E2 glycoprotein pr
42	37	46.2	520	2 S14598	E2 glycoprotein -
43	37	46.2	544	2 S41626	spike protein chai
44	37	46.2	550	1 VG1HD6	E2 glycoprotein pr
45	37	46.2	561	1 S34191	sulfite reductase

ALIGNMENTS

RESULT 1
JC7371
toxin - jellyfish (Carybdea rastoni)
C:Species: Carybdea rastoni
C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 17-Nov-2000
C:Accession: JC7371; PC7094
R:Nagai, H.; Takuwa, K.; Nakao, M.; Ito, E.; Miyake, M.; Noda, M.; Nakajima, T.
Biochem. Biophys. Res. Commun. 275, 582-588, 2000
A:Title: Novel proteinaceous toxins from the box jellyfish (sea wasp) Carybdea raston
A:Reference number: JC7371
A:Accession: JC7371
A:Molecule type: mRNA
A:Residues: 1-450 <NAG>
A:Cross-references: DDBJ:AB015878
A:Accession: PC7094
A:Molecule type: Protein
A:Residues: 39-55;56-70;196-210;250-267;268-279;309-325;363-377;378-382 <NA2>
C:Comment: This protein, a member of bioactive protein, has hemolytic activity.
C:Keywords: hemolysis; inflammation; toxin

Query Match 100.0%; Score 80; DB 2; Length 450;
Best Local Similarity 100.0%; Pred. No. 3.6e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSDGFTTMSDRRK 15
DB 363 MSDGFTTMSDRRK 377
RESULT 2
T10393
chitinase (EC 3.2.1.14), precursor - Orgyia pseudotsugata nuclear polyhedrosis virus
C:Species: Orgyia pseudotsugata nuclear polyhedrosis virus, OpNPV
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T10393
R:Ahrens, C.A.; Russell, F.R.; Funk, C.J.; Evans, J.; Harwood, S.; Rohmann, G.F.
Virology 229, 381-399, 1997
A:Title: The sequence of the Orgyia pseudotsugata multinnucleocapsid nuclear polyhedro
A:Reference number: 217011; MUID:97271300
A:Accession: T10393
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-550 <AHR>
A:Cross-references: EMBL:U75930; MID:q2934903; PIDN:AA059123.1; PID:q1911370
C:Superfamily: Serratia marcescens chitinase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
Query Match 53.8%; Score 43; DB 2; Length 550;
Best Local Similarity 46.7%; Pred. No. 14;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C/Accession: C75632
R/White, O.: Eisen, J.A.; Heidelberg, J.F.; Hickey, E.R.; Peterson, J.D.; Dodson, R.J.;
M.: Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M.
S.: Smith, H.O.; Venter, J.C.; Fraser, C.M.
science 286, 1571-1577, 1999

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-704 <TE>
A:Cross-references: GB:AE002368; GB:AE002098; NID:q7225303; PIDN:AAF40546.1; PID:q722

A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0082

Query Match 50.0%; Score 40; DB 2; Length 704;
Best Local Similarity 50.0%; Pred. No. 62;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 SDGFTYMNSDRK 15
DB 5 SDGLQSIINNRRK 18

RESULT 12

hypothetical protein 87.1k protein - bovine adenovirus 3
C:Species: Mastadenovirus bos (bovine adenovirus 3)
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 07-May-1999
C:Accession: JQ1866
R:Mittal, S.K.; Prevec, L.; Babluk, L.A.; Graham, F.L.
J. Gen. Virol. 73, 3295-3300 1992
A:Title: Sequence analysis of bovine adenovirus type 3 early region 3 and fibre protein
A:Reference number: PQ0499; MUID:93107871
A:Accession: JQ1866
A:Molecule type: DNA
A:Residues: 1-833 <MIT>
A:Cross-references: DDBJ:D12928
A:Experimental source: strain WBR-1
A:Note: The authors described carbohydrate binding site for residue 574

Query Match 50.0%; Score 40; DB 2; Length 833;
Best Local Similarity 70.0%; Pred. No. 74;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSDGFYTMEN 10
DB 104 MSDGLYTKDN 113

RESULT 13

hypothetical protein SP1831 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: G95213
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Held
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: G95213
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-160 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75904.1; PID:q14973332; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1831

Query Match 48.8%; Score 39; DB 2; Length 160;
Best Local Similarity 46.7%; Pred. No. 20;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 MSDGFYTMNSDRK 15
DB 114 MSDGFYFYONNOK 128

RESULT 14
H98077
hypothetical protein spr1650 [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: H98077

R:Hoskins, J.A.; Albhorn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.
E.R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
Y.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: H98077

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-164 <KUR>
A:Cross-references: GB:AE007317; PIDN:ALU0453.1; PID:q15459321; GSPDB:GN00174
C:Genetics:
A:Gene: spr1650

Query Match 48.8%; Score 39; DB 2; Length 164;
Best Local Similarity 46.7%; Pred. No. 20;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 MSDGFYTMNSDRK 15
DB 118 MSDGFYFYONNOK 132

RESULT 15

E84263
glycerol-3-phosphate dehydrogenase chain A [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84263
R:Ng, W.V.; Kennedy, S.P.; Mahairs, G.G.; Bergist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: E84263
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-187 <STO>
A:Cross-references: GB:AE004437; NID:q10580619; PIDN:AAG19473.1; GSPDB:GN00138
C:Genetics:
A:Gene: gpdA1

Query Match 48.8%; Score 39; DB 2; Length 187;
Best Local Similarity 46.2%; Pred. No. 23;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSDGFYTMNSDR 13
DB 103 ISRGFYVDHADR 115

Search completed: June 10, 2002, 15:22:26
Job time: 206 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 15:29:42 ; Search time 18.65 Seconds

(without alignments)
31.142 Million cell updates/sec

Title: US-09-647-522-3

Perfect score: 80

Sequence: 1 MSDFGYTMENSDRRK 15

Scoring table: BLOSUM62

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	53.8	550	1 CHIT_NPVOP	010363 oryza psu
2	42	52.5	552	1 ESTE_MYZPE	P35501 myzus persi
3	41	51.2	514	1 RS6E_SULISO	Q38046 sulfolobus
4	40	50.0	704	1 LIPA_NEIMA	P57037 neisseria m
5	40	50.0	704	1 LIPA_NEIMA	Q05013 neisseria m
6	40	50.0	976	1 FTBP_ADRE3	Q03553 bovine aden
7	39	48.8	339	1 F16O_BRANA	P46267 brassica na
8	39	48.8	357	1 YNS7_YEAST	P33873 saccharomyc
9	39	48.8	410	1 AROA_THFAC	Q91168 thermoplas
10	39	48.8	411	1 YAI2_SCHPO	Q09685 schistosac
11	39	48.8	572	1 FOXB_ECOLI	P07003 escherichia
12	38	47.5	354	1 LXA2_PHOLE	P29238 photobacter
13	38	47.5	360	1 RECF_RICPR	Q82667 rickettsia
14	38	47.5	464	1 GSHR_SCHPO	P78965 schistosac
15	38	47.5	535	1 Z257_HUMAN	Q91291 homo sapien
16	37	46.2	210	1 MOBC_BP74	P07074 bacterioph
17	37	46.2	520	1 VGL2_IBVU1	P30206 avian infec
18	37	46.2	520	1 VGL2_IBVU2	P30207 avian infec
19	37	46.2	520	1 VGL2_IBVU3	P30208 avian infec
20	37	46.2	550	1 VGL2_IBVU3	P17663 avian infec
21	37	46.2	559	1 CCSI_THI10	P52673 chloacapa r
22	37	46.2	1154	1 VGL2_IBVD2	P12722 avian infec
23	37	46.2	1162	1 VGL2_IBV4	P12223 avian infec
24	37	46.2	1162	1 VGL2_IBV4	P12650 avian infec
25	37	46.2	1163	1 VGL2_IBV6	P05133 avian infec
26	36.5	45.6	254	1 FLIP_BORBU	Q44763 borrelia bu
27	36.5	45.6	505	1 PEKA_CRIFA	P25055 citrildia f
28	36	45.0	167	1 T2DB_YEAST	P11747 saccharomyc
29	36	45.0	242	1 GLUA_CORGL	P86243 corynebacte
30	36	45.0	374	1 RECF_RHIME	P69603 rhizobium m
31	36	45.0	427	1 TRBI_AERPE	Q98855 aeropyrum p
32	36	45.0	526	1 ACHI_YEAST	P52316 saccharomyc
33	36	45.0	659	1 SYR_HUMAN	P54136 homo sapien

34	36	45.0	677	1 RN14_YEAST	P25298 saccharomyc
35	36	45.0	705	1 YASO_RHISM	P55656 rhizobium s
36	36	45.0	859	1 SYL_BUCAI	P57519 buchnera a
37	35.5	44.4	178	1 RAS_ARTSA	P18262 artemia sal
38	35	43.8	57	1 RS27_HALNL	Q9H17 halobacteri
39	35	43.8	410	1 HPLK_HAELN	P44546 haemophilus
40	35	43.8	419	1 HPLK_ECOLI	P25662 escherichia
41	35	43.8	432	1 YK27_CAEEL	O16686 caenorhabdi
42	35	43.8	463	1 DNMA_RICPR	O59758 rickettsia
43	35	43.8	539	1 ERG1_PANGI	O48651 panax gins
44	35	43.8	545	1 CIP4_HUMAN	O15642 homo sapien
45	35	43.8	746	1 NUSC_VICFA	P15958 victa faba

ALIGNMENTS

RESULT	ID	CHIT_NPVOP	STANDARD	PRT	550 AA.
AC	010363				
DT	01-NOV-1997	(Rel. 35, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DT	01-NOV-1997	(Rel. 35, Last annotation update)			
DE	Probable endochitinase precursor (EC 3.2.1.14).				
OS	Oryza pseudotsugata multicapsid polynucleoside virus (OPMPV).				
OC	Viruses; dsDNA viruses, no RNA stage; Baculoviridae;				
OX	Nucleopolyhedrovirus.				
NCBI_Taxid=164623;					
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-97271300; PubMed-9126251;				
RA	Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,				
RA	Rohmann G.F.;				
RT	"The sequence of the Oryza pseudotsugata multnucleosid nuclear				
RT	polynucleoside virus genome."				
RL	Virology 229:381-399(1997).				
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-				
CC	acetyl-D-glucosamine polymers of chitin.				
CC	-1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (Potential).				
CC	-1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL				
CC	HYDROLASES).				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
DR	EMBL: U75930; AAC59123.1; -				
DR	HSSP: P07254; ICTN				
DR	InterPro: IPR001579; Chitinase_2.				
DR	InterPro: IPR000886; ER_target.				
DR	InterPro: IPR001223; Glyco_hydro_18.				
DR	InterPro: IPR000601; PKD_domain.				
DR	Pfam: PF00704; Glyco_hydro_18; 1.				
DR	SMART: SM00089; PKD; 1.				
DR	PROSITE: PS00014; ER_TARGET; 1.				
DR	PROSITE: PS01095; CHITINASE_18; 1.				
KW	Hydrolase; Glycosidase; Chitin degradation; Signal; Glycoprotein;				
KW	Endoplasmic reticulum.				
FT	STGNL	1	16		
FT	CHAIN	17	550		POTENTIAL.
FT	ACT_SITE	304	304		PROBABLE ENDOCHITINASE.
FT	CARBOHYD	146	146		PROTON DONOR (BY SIMILARITY).
FT	CARBOHYD	172	172		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	344	344		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	STATE	547	550		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SEQUENCE	550 AA;	60733 MW;	77947FSCF00E07BD CRC64;	PREVENT SECRETION FROM ER (POTENTIAL).

Query Match 53.8%; Score 43; DB 1; Length 550;
 Best Local Similarity 46.7%; Pred. No. 5;
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSGFTTENSRRK 15
 :|||:|:|:|:
 DB 266 LSDPFYDNEBRR 280

RESULT 2
 ESTE_MYZPE STANDARD; PRT; 552 AA.
 ID ESTE_MYZPE P35501;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE Esterase E4 precursor (Ec 3.1.1.1) (Carboxylic-ester hydrolase).
 OS Myzus persicae (Peach-potato aphid).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha;
 OC Aphidiformes; Aphidoidea; Aphididae; Macrosiphini; Myzus.
 NCBI_TaxID=13164;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-63.
 RC STRAIN-R3 / ISOLATE 794J;
 RX MEDLINE=93384534; PubMed=8373371;
 RA Field L.M., Williamson M.S., Moores G.D., Devonshire A.L.;
 RT "Cloning and analysis of the esterase genes conferring insecticide
 resistance in the peach-potato aphid, Myzus persicae (Suizet).";
 RL Biochem. J. 294:569-574(1993).
 CC -1 FUNCTION: OVERPRODUCTION OF NONSPECIFIC ESTERASES IS A COMMON
 MECHANISM OF RESISTANCE TO ORGANOPHOSPHATE INSECTICIDES.
 CC -1 CATALYTIC ACTIVITY: A carboxylic ester + H(2)O - an alcohol + a
 carboxylic anion.
 CC -1 MISCELLANEOUS: THIS ESTERASE CONFERS INSECTICIDE RESISTANCE.
 CC -1 SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X74554; CAAS2648.1;
 DR PIR: S36786; S36786.
 DR HSSP: P21836; 1MAH.
 DR InterPro: IPR002018; Carboxylesterase-B.
 DR Pfam: PF00135; Coesterase; Est_1lp_thioest_actsite.
 DR PROSITE: PS00122; CARBOXYLESTERASE_B_1; 1.
 DR PROSITE: PS00941; CARBOXYLESTERASE_B_2; FALSE_NEG.
 KW Hydrolyase; Serine esterase; Glycoprotein; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 552
 FT ACT_SITE 214 214 ESTERASE E4.
 FT ACT_SITE 339 339 BY SIMILARITY.
 FT ACT_SITE 463 463 BY SIMILARITY.
 FT DISULFID 89 106 BY SIMILARITY.
 FT DISULFID 266 277 BY SIMILARITY.
 FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 371 371 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SEQUENCE 552 AA; 61348 MW; B97B67272DFE7209 CRC64;

Query Match 52.5%; Score 42; DB 1; Length 552;
 Best Local Similarity 53.8%; Pred. No. 7.5;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 DGFYDNEBRRK 15
 :|||:|:|:|:
 DB 474 DGFYDNEBRRK 486

RESULT 3
 RS6E_SULSO STANDARD; PRT; 214 AA.
 ID RS6E_SULSO O980A6;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE 30S ribosomal protein S6e.
 GN RS6E OR SS00411.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; PubMed=11427726;
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Aweez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moors A., Erasmo G., Fletcher C., Gordon P.M.K.,
 RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 CC -1 SIMILARITY: BELONGS TO THE S6E FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AE006674; AAK40739.1;
 DR InterPro: IPR001377; Ribosomal_S6E.
 DR Pfam: PF01092; Ribosomal_S6e; 1.
 DR ProDom: PD003460; Ribosomal_S6E; 1.
 DR PROSITE: PS00578; RIBOSOMAL_S6E; 1.
 DR Ribosomal protein; Complete proteome.
 KW Ribosomal protein; Complete proteome.
 SEQUENCE 214 AA; 23720 MW; 9E20B93EB0247693 CRC64;

Query Match 51.2%; Score 41; DB 1; Length 214;
 Best Local Similarity 58.3%; Pred. No. 4.1;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 GFTYDNEBRRK 15
 :|||:|:|:|:
 DB 182 GFTYDNEBRRK 193

RESULT 4
 LIPA_NEIMA STANDARD; PRT; 704 AA.
 ID LIPA_NEIMA P57037;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Capsule polysaccharide modification protein lpa.
 GN LIPA OR NMA0186.
 GN Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 NCBI_TaxID=65699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;

RA MEDLINE-20222556; PubMed-10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davis R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
 RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrall B.G.;
 RA "Complete DNA sequence of a serogroup A strain of *Neisseria*
 RT meningitidis 22491.";
 RL Nature 404:502-506(2000).
 CC -1- FUNCTION: INVOLVED IN THE PHOSPHOLIPID MODIFICATION OF THE
 CC CAPSULAR POLYSACCHARIDE. A STRONG REQUIREMENT FOR ITS
 CC TRANSLLOCATION TO THE CELL SURFACE.
 CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE-ASSOCIATED (CYTOPLASMIC
 CC SIDE) (PROBABLE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL162752; CAB83501.1;
 DR Inner membrane; Polysaccharide transport; Transport;
 KM Complete proteome.
 SO SEQUENCE 704 AA; 79565 MW; 2E1C5D665D9B861 CRC64;
 QY 2 SDGFTYMENSDDRK 15
 DB 5 SDGLQSIINNRRK 18
 Query Match 50.0%; Score 40; DB 1; Length 704;
 Best Local Similarity 50.0%; Pred. No. 22;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

RESULT 5
 LIPA_NEIMB STANDARD; PRT; 704 AA.
 AC 005013;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Capsule polysaccharide modification protein lIpa.
 GN LIPA OR NMB0082.
 OS *Neisseria meningitidis* (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; *Neisseria*.
 OX NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / SEROGROUP B.
 RX MEDLINE-20175755; PubMed-10710307;
 RA Tettein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 RA Nelson W.C., Gwyn M.L., Debey R., Peterson J.D., Hickey E.K.,
 RA Halt D.H., Salzberg S.L., White O., Fleischmann K.D., Dougherty B.A.,
 RA Mason T., Clecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
 RA Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J.,
 RA Gill J., Scariato V., Masignani V., Pizze M., Grandi G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
 RT "Complete genome sequence of *Neisseria meningitidis* serogroup B strain
 RT MC58.";
 RL Science 287:1809-1815(2000).
 RN [2]
 RP SEQUENCE OF 184-704 FROM N.A.
 RC STRAIN=B1940 / SEROGROUP B;
 RX MEDLINE-93316845; PubMed-8326861;
 RA Frosch M., Mueller A.;
 RT "Phospholipid substitution of capsular polysaccharides and mechanisms
 RT of capsule formation in *Neisseria meningitidis*.";

RL MOL. Microbiol. 8:483-493(1993).
 CC -1- FUNCTION: INVOLVED IN THE PHOSPHOLIPID MODIFICATION OF THE
 CC CAPSULAR POLYSACCHARIDE. A STRONG REQUIREMENT FOR ITS
 CC TRANSLLOCATION TO THE CELL SURFACE.
 CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE-ASSOCIATED (CYTOPLASMIC
 CC SIDE) (PROBABLE).
 CC -1- CAUTION: REP. 2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION
 CC 566 ONWARD DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE002367; AAP40546.1;
 DR EMBL: Z13995; NOT_ANNOTATED_CDS.
 DR PIR: S28077; S28077.
 DR PIR: S32879; S32879.
 DR TIGR: NMB0082;
 KM Inner membrane; Polysaccharide transport; Transport;
 KM Complete proteome.
 FT CONFLICT 238
 FT CONFLICT 253
 FT CONFLICT 293
 FT CONFLICT 294
 FT CONFLICT 306
 FT CONFLICT 316
 FT CONFLICT 331
 FT CONFLICT 330
 FT CONFLICT 390
 FT CONFLICT 449
 FT CONFLICT 456
 FT CONFLICT 462
 SO SEQUENCE 704 AA; 79605 MW; 2909C40642CD326A CRC64;
 QY 2 SDGFTYMENSDDRK 15
 DB 5 SDGLQSIINNRRK 18
 Query Match 50.0%; Score 40; DB 1; Length 704;
 Best Local Similarity 50.0%; Pred. No. 22;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

RESULT 6
 FIBP_ADEB3 STANDARD; PRT; 976 AA.
 AC 003553;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Fiber protein.
 GN FIV.
 OS Bovine adenovirus type 3 (Mastadenovirus bos3).
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 OX NCBI_TaxID=10510;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93107871; PubMed-1469367;
 RA Mittal S.K., Prevec L., Babluk L.A., Graham F.L.;
 RT "Sequence analysis of bovine adenovirus type 3 early region 3 and
 RT fibre protein genes.";
 RL J. Gen. Virol. 73:3295-3300(1992).
 RN [2]
 RP REVISIONS.
 RX MEDLINE-94103794; PubMed-8277294;
 RA Mittal S.K., Prevec L., Babluk L.A., Graham F.L.;
 RT "Sequence analysis of bovine adenovirus type 3 early region 3 and
 RT fibre protein genes.";
 RL J. Gen. Virol. 74:2825-2825(1993).
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN-MBR-1;
 RX MEDLINE-98105785; PubMed-9445040;
 RA Reddy P.S., Idamakanti N., Zakharchouk A.N., Baxi M.K., Lee J.B.,
 RA Pyne C., Babluk L.A., Tikoo S.K.;
 RT "Nucleotide sequence, genome organization, and transcription map of
 RT bovine adenovirus type 3.";
 RL J. Virol. 72:1394-1402(1998).
 CC -1- FUNCTION: RECOGNIZES THE CELL RECEPTOR; SERVES AS THE LIGAND
 CC BETWEEN THE ADENOVIRUS CAPSID AND THE HOST CELL RECEPTOR.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D16839; BAA04115.1; -.
 DR EMBL: AF030154; AAD09736.1; -.
 DR InterPro: IPR000939; Adeno_fiber2.
 DR InterPro: IPR000978; Adeno_fiber_knob.
 DR InterPro: IPR000931; Adeno_fibre.
 DR Pfam: PF00541; adeno_fiber; 1.
 DR Pfam: PF00608; adeno_fiber2; 16.
 DR PRINTS: PR00307; ADENOVSFIBRE.
 KM Fiber protein.
 SO SEQUENCE 976 AA; 102323 MW; 9D0EC36052F02896 CRC64;

Query Match 50.0%; Score 40; DB 1; Length 976;
 Best Local Similarity 70.0%; Pred. No. 31;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 MSDGYTMEN 10
 DB 97 MSDGYTKDN 106

RESULT 7
 ID F160_BRANA STANDARD; PRT; 339 AA.
 AC P46267;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Fructose-1,6-bisphosphatase, cytosolic (EC 3.1.3.11) (D-fructose-1,6-
 DE bisphosphate 1-phosphohydrolase) (FBPase).
 OS Brassica napus (Rape).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Brassica.
 CC NCB1_TaxID=3708;
 RN NCB1 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cotyledon;
 RX MEDLINE-95537438; PubMed-7630967;
 RA Laroche A., Fick M.M., Kazala C., Weselake R.J., Thomas J.E.;
 RT "Isolation and characterization of an oilseed rape fructose-1,6-
 RT bisphosphatase cDNA.";
 RL Plant Physiol. 108:1335-1336(1995).
 CC -1- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate + H(2)O -> D-
 CC fructose 6-phosphate + phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- MISCELLANEOUS: IN PLANTS THERE ARE TWO FBPASE ISOZYMES: ONE IN THE
 CC CYTOSOL AND THE OTHER IN THE CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE FBPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U20179; AAA82750.1; -.
 DR HSSP: P0637; 1BK4.
 DR InterPro: IPR000146; In_FB_phphatase.
 DR Pfam: PF00316; FBPase; 1.
 DR PRINTS: PR00377; INFBPHPTASE.
 DR PRODOM: PD001491; In_FB_phphatase; 1.
 DR PROSITE: PS00124; FBPASE; 1.
 KM Hydroxylase: Carbohydrate metabolism; Multigene family.
 FT ACT SITE 276 BY SIMILARITY
 SO SEQUENCE 339 AA; 37156 MW; 3844CD90F3C6DD33 CRC64;

Query Match 48.8%; Score 39; DB 1; Length 339;
 Best Local Similarity 77.8%; Pred. No. 15;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 GFTYMNSD 12
 DB 139 GFTYMNSD 147

RESULT 8
 ID YNS7_YEAST STANDARD; PRT; 357 AA.
 AC P53873;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Hypothetical 40.3 kDa protein in KAR1-UBP10 intergenic region.
 DE YNL187W OR N1615.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 CC NCB1_TaxID=4932;
 RN NCB1 [1]
 RP SEQUENCE FROM N.A.
 RA Obermayer B., Piravandi E., Rinke M., Domdey H.;
 RL submitted (May-1996) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z71463; CA96081.1; -.
 DR SGD: S0005131; YNL187W.
 KM Hypothetical protein.
 SO SEQUENCE 357 AA; 40307 MW; ED52B00C4902453A CRC64;

Query Match 48.8%; Score 39; DB 1; Length 357;
 Best Local Similarity 69.2%; Pred. No. 16;
 Matches 9; Conservative 1; Mismatches 1; Indels 2; Gaps 1;

OY 3 DGYTYM--ENSDR 13
 DB 65 DGYTYMNSDR 77

RESULT 9
 ID AROA_THEAC STANDARD; PRT; 410 AA.
 AC Q9HUE6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

1
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19)
GN AROA OR TA0282.
OS Thermoplasma acidophilum.
OC Archaea: Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;
OC Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
  Mewes H.-W., Fishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermophilic scavenger Thermoplasma
  acidophilum."
RT Nature 407:508-513(2000).
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
  phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -1- PATHWAY: SIXTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN THE
  BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE EPSP SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL445063; CAC11427.1;
DR InterPro: IPR001986; EPSP_synthase.
DR Pfam: PF00275; EPSP_synthase.1.
DR ProDom: PD001867; EPSP_synthase.1.
DR PROSITE: PS00104; EPSP_SYNTHASE_1; FALSE_NEG.
DR PROSITE: PS00885; EPSP_SYNTHASE_2; FALSE_NEG.
KW Aromatic amino acid biosynthesis; Transferrase; Complete proteome.
SQ SEQUENCE 410 AA; 4516 MW; 0EF167698CC2FA26 CRC64;

Query Match 48.8%; Score 39; DB 1; Length 410;
Best Local Similarity 46.2%; Pred. No. 18;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 DGEYTMENSDRRK 15
Db 136 DGEYDVGSESK 148

RESULT 10
YAL2_SCHPO STANDARD; PRT; 411 AA.
ID YAL2_SCHPO
AC 009685;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 48.5 kDa protein C13C5.02 in chromosome I.
GN SPAC13C5.02.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
  Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 WW DOMAINS.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z50112; CA90453.1;
DR InterPro: IPR002713; FE.
DR InterPro: IPR001202; WW.
DR InterPro: IPR002349; WW_domain.
DR Pfam: PF01846; FE; 1.
DR Pfam: PF00397; WW; 2.
DR PRINTS: PR00403; WMDOMAIN.
DR SMART: SM00441; FE; 1.
DR SMART: SM00456; WW; 2.
DR PROSITE: PS01159; WW_DOMAIN_1; 2.
DR PROSITE: PS50020; WW_DOMAIN_2; 2.
KW Hypothetical protein; Repeat.
FT DOMAIN 3 36 WW 1.
FT DOMAIN 89 122 WW 2.
SQ SEQUENCE 411 AA; 48519 MW; 7A89C5F7397B8A8F CRC64;

Query Match 48.8%; Score 39; DB 1; Length 411;
Best Local Similarity 46.2%; Pred. No. 19;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 DGEYTMENSDRRK 15
Db 272 DREYVLDSEGRKK 284

RESULT 11
POXB_ECOLI STANDARD; PRT; 572 AA.
ID POXB_ECOLI
AC P07003; Q47513; Q47514; Q47515; Q47516; Q47517; Q47518; Q47519;
AC Q47520;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pyruvate dehydrogenase [cytochrome] (EC 1.2.2.2) (Pyruvate oxidase)
DE (POX) (Pyruvate dehydrogenase [ubiquinone]).
GN POXB OR B0871.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-K12;
RX MEDLINE=86286555; PubMed=3016647;
RX Grabau C., Cronan J.E. Jr.;
RT "Nucleotide sequence and deduced amino acid sequence of Escherichia
  coli pyruvate oxidase, a lipid-activated flavoprotein."
RL Nucleic Acids Res. 14:5449-5460(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97426617; PubMed=9278503;
RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
  Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
  Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
  Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97061202; PubMed=8905232;
RX Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
  Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,

```

RA Kimura S., Kitaagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 [4]
 RN SEQUENCE FROM N.A., AND MUTAGENESIS.
 RC STRAIN-K12;
 RX MEDLINE-89308683; PubMed-2663858;
 RA Graba C., Chang Y.Y., Cronan J.E. Jr.;
 RT "Lipid binding by *Escherichia coli* pyruvate oxidase is disrupted by
 RT small alterations of the carboxyl-terminal region.";
 RL J. Biol. Chem. 264:12510-12519(1989).
 [5]
 RN SEQUENCE OF 550-572 FROM N.A.
 RP MEDLINE-86033917; PubMed-3902830;
 RA Rechy M.A., Graba C., Cronan J.E. Jr., Hager L.P.;
 RT "Characterization of the alpha-peptide released upon protease
 RT activation of pyruvate oxidase.";
 RL J. Biol. Chem. 260:14287-14291(1985).
 [6]
 RP SEQUENCE OF 1-22 FROM N.A.
 RX MEDLINE-94293772; PubMed-8022274;
 RA Chang Y.Y., Wang A.Y., Cronan J.E. Jr.;
 RT "Expression of *Escherichia coli* pyruvate oxidase (Pox) depends on
 RT the sigma factor encoded by the *proS(katF)* gene.";
 RL Mol. Microbiol. 11:1019-1028(1994).
 CC -1- CATALYTIC ACTIVITY: Pyruvate + ferricytochrome b1 + H(2)O = CO(2)
 CC + acetate + ferrocyanochrome b1.
 CC -1- COFACTOR: THIAMINE PYROPHOSPHATE, FAD AND MAGNESIUM.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- PTM: ACTIVATED BY LIMITED PROTEOLYTIC DIGESTION. THIS CLEAVAGE
 CC PRODUCES A PEPTIDE (ALPHA-PEPTIDE) AND MIMICS THE ACTIVATION OF
 CC ENZYME BY PHOSPHOLIPIDS. THE PROTEOLYTIC CLEAVAGE ALSO RESULTS IN
 CC THE LOSS OF THE HIGH AFFINITY LIPID-BINDING SITE OF THE ENZYME.
 CC -1- SIMILARITY: WITH OTHER ENZYMES WHICH REQUIRE TPP.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X04105; CAA27725.1; -
 DR EMBL: AE000188; AAC73958.1; -
 DR EMBL: D90724; BAA35585.1; -
 DR EMBL: S73268; AAB31180.1; -
 DR EMBL: M28208; AAB59101.1; -
 DR EMBL: L47688; AAB59102.1; -
 DR EMBL: L47689; AAB59103.1; -
 DR EMBL: L47690; AAB59104.1; -
 DR EMBL: L47691; AAB59105.1; -
 DR EMBL: L47692; AAB59106.1; -
 DR EMBL: L47693; AAB59107.1; -
 DR EMBL: L47694; AAB59108.1; -
 DR EMBL: L47695; AAB59109.1; -
 DR PIR: A23648; DEECPC.
 DR HSSP: P37063; 1POM.
 DR SWISS-2DPAGE: P07003; COLI.
 DR ECODBASE: G058.0; 6TH EDITION.
 DR ECOGENE: EG10754; poxb.
 DR InterPro: IPR000399; TPP_enzyme.
 DR Pfam: PF00205; TPP_enzymes_1.
 DR Pfam: PF02775; TPP_enzymes_C_1.
 DR Pfam: PF02776; TPP_enzymes_N_1.
 DR PROSITE: PS00187; TPP_ENZYMES; 1.
 KW Oxidoreductase; Flavoprotein; FAD; Thiamine pyrophosphate; Magnesium;
 Membrane; Lipid-binding; Complete proteome.

FT PEPTIDE 550 572 ALPHA-PEPTIDE.
 FT ACT SITE 50 50 BY SIMILARITY.
 FT MUTAGEN 533 533 A->T: IN POXB11.
 FT MUTAGEN 533 533 A->V: IN POXB14.
 FT MUTAGEN 560 560 D->P: IN POXB15; NORMAL ACTIVITY.
 FT MUTAGEN 564 564 E->P: IN POXB16; LOSS OF ACTIVITY.
 FT MUTAGEN 572 572 R->G: IN POXB10; REDUCED ACTIVITY; MAY
 FT INTERACT LESS WITH MEMBRANES.
 FT MUTAGEN 549 572 MISSING: IN POXB6.
 FT MUTAGEN 564 572 MISSING: IN POXB7.
 FT MUTAGEN 570 572 MISSING: IN POXB8.
 FT CONFLICT 364 365 QQ -> HE (IN REF. 4).
 FT CONFLICT 414 416 QAL -> HGV (IN REF. 4).
 SQ SEQUENCE 572 AA; 62011 MW; 57B38B9E3A92BDA CRC64;
 Query Match 48.8%; Score 39; DB 1; Length 572;
 Best Local Similarity 63.6%; Pred. No. 26;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Oy 2 SDGYTMNSD 12
 Db 257 SSGFTMMNAD 267
 RESULT 12
 LXA2_PROLE
 ID LXA2_PROLE STANDARD; PRT; 354 AA.
 AC P29238;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alkalal monooxygenase alpha chain (EC 1.14.14.3) (Bacterial luciferase
 DE alpha chain).
 GN LUXA.
 OS Photobacterium leiognathi.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae;
 OC Photobacterium.
 OX NCBI_TaxID:658;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 25521;
 RX MEDLINE-92007870; PubMed-1915359;
 RA Lee C.Y., Szttnier R.B., Meighen E.A.;
 RT "The lux genes of the luminous bacterial symbiont, *Photobacterium*
 RT leiognathi, of the porphyry. Nucleotide sequence, difference in gene
 RT organization, and high expression in mutant *Escherichia coli*.";
 RL Eur. J. Biochem. 201:161-167(1991).
 CC -1- FUNCTION: LIGHT-EMITTING REACTION IN LUMINOUS BACTERIA.
 CC -1- CATALYTIC ACTIVITY: RCHO + FMN(H2) + O(2) = RCOOH + FMN + H(2)O +
 CC light.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M63594; AAA25618.1; -
 DR PIR: S17953; S17953.
 DR HSSP: P07740; 1LUC.
 DR InterPro: IPR002103; Bac_luciferase.
 DR Pfam: PF00296; bac_luciferase; 1.
 DR PRINTS: PR00089; LUCIFERASE
 DR PROSITE: PS00494; BACTERIAL_LUCIFERASE; 1.
 KW Photoprotein; Luminescence; Oxidoreductase; Monooxygenase;
 FMN
 FT DOMAIN 100 115 REGION OF ACTIVE CENTER-1 (BY
 FT SIMILARITY).
 FT DOMAIN 279 294 REGION OF ACTIVE CENTER-2 (BY
 FT SIMILARITY).

FT SEQUENCE 354 AA; 40402 MW; 07001C2064F4645B CRC64;
 SQ SIMILARITY)
 Query Match 47.5%; Score 38; DB 1; Length 354;
 Best Local Similarity 75.0%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 DGEYTMEN 10
 Db 37 DGEYTMEN 44

RESULT 13
 RECF_RICPR STANDARD; PRT; 360 AA.
 AC Q2EB6;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA replication and repair protein recf.
 GN RECF OR RP029.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 RX NCBI_Taxid:782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MADRID E;
 RA MEDLINE-99039499; PubMed-9823893;
 RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
 RA Skerhelt-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 mitochondria.";
 RL Nature 396:133-140(1998).
 CC -1- FUNCTION: THE RECF PROTEIN IS INVOLVED IN DNA METABOLISM; IT IS
 CC REQUIRED FOR DNA REPLICATION AND NORMAL SOS INDICIBILITY. RECF
 CC BINDS PREFERENTIALLY TO SINGLE-STRANDED, LINEAR DNA. IT ALSO SEEMS
 CC TO BIND ATP (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO THE RECF FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AJ235270; CAI4500.1; -
 DR InterPro: IPR001238; RECF.
 DR Pfam: PF00470; RECF.1.
 DR PROSITE: PS00617; RECF_1; 1.
 DR PROSITE: PS00618; RECF_2; 1.
 KW DNA damage; DNA replication; DNA-binding; SOS response; DNA repair;
 KM ATP-binding; Complete proteome.
 FT NP_BIND 33 40 ATP (POTENTIAL).
 SO SEQUENCE 360 AA; 41489 MW; 500B37B8BD77A16D CRC64;

Query Match 47.5%; Score 38; DB 1; Length 360;
 Best Local Similarity 46.2%; Pred. No. 24;
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 DGEYTMENSDRRK 15
 Db 125 EGIPTSSSTDRK 137

RESULT 14
 GSHR_SCHPO

ID GSHR_SCHPO STANDARD; PRT; 464 AA.
 AC P78965; O13631;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Glutathione reductase (EC 1.6.4.2) (Gr) (Grase).
 GN PGRI OR SPBC17A3.07 OR P1039.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 RX NCBI_Taxid=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RX MEDLINE-97435263; PubMed-9287302;
 RA Lee J., Daves I.W., Roe J.H.;
 RT "Isolation, expression, and regulation of the pgri(+) gene encoding
 RT glutathione reductase absolutely required for the growth of
 RT Schizosaccharomyces pombe.";
 RL J. Biol. Chem. 272:23042-23049(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA Kushida N., Yamazaki S., Tanaka T., Jinno K., Halkawa Y., Yamazaki J.,
 RA Yamamoto S., Sekine M., Oguchi A., Nagai Y., Sakai M., Aoki K.,
 RA Ogura K., Otsuka R., Kudoh Y., Yanagida M., Machida M., Zhang M.O.;
 RL Submitted (May-1997) to the EMBL/GenBank/DBD databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA Wood V., Skelton J., Churcher C.M., Rajandream M.A., Barrell B.G.;
 RL Submitted (Jul-1999) to the EMBL/GenBank/DBD databases.
 CC -1- FUNCTION: MAINTAIN HIGH LEVELS OF REDUCED GLUTATHIONE IN THE
 CC CYTOSOL.
 CC -1- CATALYTIC ACTIVITY: NADPH + oxidized glutathione = NADP(+) + 2
 CC glutathione.
 CC -1- COFACTOR: FAD (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
 CC OXIDOREDUCTASES CLASS-I.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U63845; AAC49809.1; -
 DR EMBL: AB004537; BAA21419.1; -
 DR EMBL: AL109652; CAB51766.1; -
 DR HSP: P00390; IALG.
 DR InterPro: IPR001327; FAD_Pyr_redox.
 DR InterPro: IPR000815; Hg_reductase.
 DR InterPro: IPR000103; Pyridine_redox_2.
 DR InterPro: IPR001100; Pyr_redox.
 DR InterPro: IPR004099; Pyr_redox_dlm.
 DR Pfam: PF00070; Pyr_redox_1.
 DR Pfam: PF02852; Pyr_redox_dlm_1.
 DR PRINTS: PR00368; FADPFR.
 DR PRINTS: PR00945; HGRDTASE.
 DR PRINTS: PR00411; PNDRTASEI.
 DR PRINTS: PR00469; PNDRTASEII.
 DR PROSITE: PS00076; PYRIDINE_REDOX_1; 1.
 KW Redox-active center; Oxidoreductase; Flavoprotein; FAD; NADP.
 FT NP_BIND 9 39
 FT DISULFID 46 51 REDOX-ACTIVE (BY SIMILARITY).
 FT NP_BIND 305 315 FAD (FLAVIN PART) (BY SIMILARITY).
 FT ACT_SITE 453 453 BY SIMILARITY.

FT CONFLICT 184 184 V -> VV (IN REF. 1).
 FT CONFLICT 419 424 LHLVGD -> PPFSSWR (IN REF. 1).
 SQ SEQUENCE 464 AA; 49999 MW; 2BFEPED363A3F173 CRC64;

Query Match 47.5%; Score 38; DB 1; Length 464;
 Best Local Similarity 46.2%; Pred. No. 32;
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 SDGFYFMSDRR 14
 Db 168 SDGFYFMSDRR 180

RESULT 15
 ID 2257 HUMAN STANDARD; PRT; 535 AA.

AC 09Y201;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Zinc finger protein 257 (Bone marrow zinc finger 4) (BMZF-4).
 GN ZNF257 OR BMZF4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.
 RC TISSUE=bone marrow;
 RX MEDLINE=20054457; PubMed=10585455;
 RA Han Z.-G., Zhang O.-H., Ye M., Kan L.-X., Gu B.-W., He K.-L.,
 RA Shi S.-L., Zhou J., Fu G., Mao M., Chen S.-J., Yu L., Chen Z.;
 RT "Molecular cloning of six novel Kruppel-like zinc finger genes from
 RT hematopoietic cells and identification of a novel transregulatory
 RT domain KRNB.";
 RL J. Biol. Chem. 274:35741-35748(1999).
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: AF070651; AAD20957.1; -
 DR HSSP: P08047; ISP2.
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; Znf-C2H2.
 DR Pfam: PF01352; KRAB; 1.
 DR Pfam: PF00096; zf-C2H2; 10.
 DR PRINTS: PRO0048; ZINCINGER.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00355; Znf-C2H2; 10.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 10.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 12.
 DR Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KM Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KW Nuclear protein; Repeat.

FT DOMAIN 4 75 KRAB.
 FT DOMAIN 173 509 ZINC_FINGERS.
 FT ZN_FING 173 195 C2H2-TYPE.
 FT ZN_FING 201 223 C2H2-TYPE.
 FT ZN_FING 229 251 C2H2-TYPE.
 FT ZN_FING 257 282 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 288 310 C2H2-TYPE.
 FT ZN_FING 316 338 C2H2-TYPE.

FT ZN_FING 344 366 C2H2-TYPE.
 FT ZN_FING 372 397 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 403 425 C2H2-TYPE.
 FT ZN_FING 431 453 C2H2-TYPE.
 FT ZN_FING 459 481 C2H2-TYPE.
 FT ZN_FING 487 509 C2H2-TYPE.
 SQ SEQUENCE 535 AA; 62348 MW; 22DC580C4613BC51 CRC64;

Query Match 47.5%; Score 38; DB 1; Length 535;
 Best Local Similarity 63.6%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 5 FYFMSDRR 15
 Db 154 FYFMSDRR 164

Search completed: June 10, 2002, 15:29:43
 Job time: 478 sec

DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
 DE GROWTH HORMONE RECEPTOR.
 GN GHR.
 OS Carassius auratus (Goldfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Carassius.
 NCBI_TaxID=7957;
 RX NCBI_TaxID=7957;
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21297186; PubMed=11404009;
 RX Lee L.T.O., Nong G., Chan Y.H., Tse D.L.Y., Cheng C.H.K.;
 RT "Molecular cloning of a teleost growth hormone receptor and its
 RT functional interaction with human growth hormone.";
 RL Gene 270:121-129(2001).
 DR EMBL; AF293417; AAK60495.1;
 KW Receptor.
 SO SEQUENCE 602 AA; 67262 MW; 4D9F66821C1A029E CRC64;

Query Match 52.5%; Score 42; DB 13; Length 602;
 Best Local Similarity 57.1%; Pred. No. 43;
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSDGYTMENSDRR 14
 Db 476 VSDGAVTSENTARQ 489

RESULT 3
 OGLIMS
 ID OGLIMS PRELIMINARY; PRT; 684 AA.
 AC OGLIMS;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
 DE GENOMIC DNA, CHROMOSOME 3, BAC CLONE:FB12.
 OS Arabidopsis thaliana (mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI_TaxID=3702;
 RX NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RA STRAIN=COLUMBIA;
 RX MEDLINE=20363099; PubMed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl.
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 DR EMBL; AP001299; BAB02564.1;
 DR EMBL; AP000370; BAB02564.1; JOINED.
 SO SEQUENCE 684 AA; 75577 MW; 6A4714690761B66A CRC64;

Query Match 52.5%; Score 42; DB 10; Length 684;
 Best Local Similarity 46.7%; Pred. No. 49;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 MSDGYTMENSDRR 15
 Db 329 LNDGYTTEEDERR 343

RESULT 4
 O980A6

ID O980A6 PRELIMINARY; PRT; 214 AA.
 AC O980A6;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, last sequence update)
 DT 01-OCT-2001 (TREMBlrel. 18, last annotation update)
 DE SSU RIBOSOMAL PROTEIN S6E (RPS6E).
 GN RPS6E.
 OS Sulfolobus solfataricus.
 CC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 NCBI_TaxID=2287;
 RX NCBI_TaxID=2287;
 RP SEQUENCE FROM N.A.
 RA STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; PubMed=11427726;
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Aveyez M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moers A., Erasmo G., Fletcher C., Gordon P.M.K.,
 RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doollittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 DR EMBL; AE006674; AAK40739.1;
 DR InterPro: IPR001377; Ribosomal_S6E.
 DR Pfam: PF01092; Ribosomal_S6E.
 DR ProDom: PD003460; Ribosomal_S6E; 1.
 KW Ribosomal protein; Complete proteome.
 SO SEQUENCE 214 AA; 23720 MW; 9E20B93EB0247693 CRC64;

Query Match 51.2%; Score 41; DB 17; Length 214;
 Best Local Similarity 58.3%; Pred. No. 21;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 4 GGYTMENSDRR 15
 Db 182 GGYTMENSDRR 193

RESULT 5
 O45606
 ID O45606 PRELIMINARY; PRT; 387 AA.
 AC O45606;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, last annotation update)
 DE FUNCTION UNKNOWN.
 GN YSCP.
 OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 NCBI_TaxID=1423;
 RX NCBI_TaxID=1423;
 RP SEQUENCE FROM N.A.
 RA STRAIN=168;
 RX MEDLINE=96051385; PubMed=7584024;
 RA Ogasawara N., Nakai S., Yoshikawa H.;
 RT "Systematic sequencing of the 180 kilobase region of the Bacillus
 RT subtilis chromosome containing the replication origin.";
 RL DNA Res. 1:1-14(1994).
 RN [2]

Query Match 52.5%; Score 42; DB 10; Length 684;
 Best Local Similarity 46.7%; Pred. No. 49;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 MSDGYTMENSDRR 15
 Db 329 LNDGYTTEEDERR 343

RESULT 4
 O980A6

RA Calogero S., Gardan R., Glaser P., Schweizer J., Rapoport G.,
 RA Debarbouille M.;
 RT "RocR, a novel regulatory protein controlling arginine utilization in
 RT *Bacillus subtilis*, belongs to the NtrC/NtrA family of transcriptional
 RT activators.";
 RL J. Bacteriol. 176:1234-1241(1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE=94236234; PubMed=8180695;
 RA Hartford O.M., Dows B.C.;
 RT "Isolation and characterization of a hydrogen peroxide resistant
 RT mutant of *Bacillus subtilis*.";
 RL Microbiology 140:297-304(1994).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE=96093926; PubMed=7584049;
 RA Yoshida K., Seki S., Fujimura M., Miwa Y., Fujita Y.;
 RT "Cloning and sequencing of a 36-kb region of the *Bacillus subtilis*
 RT genome between the *gnt* and *iol* operons.";
 RL DNA Res. 2:61-69(1995).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE=95311309; PubMed=7540694;
 RA Gardan R., Rapoport G., Debarbouille M.;
 RT "Expression of the *rocDEF* operon involved in arginine catabolism in
 RT *Bacillus subtilis*.";
 RL J. Mol. Biol. 249:843-856(1995).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Berrero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Boriss R., Boursier L., Brans A., Braun M., Bridgell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denliot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Ertlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Frits C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Gilm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi C.,
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,
 RA Prescan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadleir Y.,
 RA Sato T., Scanlan E., Schleicher S., Schroeter R., Scofield F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Solio B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemori A.,
 RA Takeuchi M., Takemori A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambut R., Wedler H., Wedler H., Weisenegger T.,
 RA Winters P., Wipst A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the gram-positive bacterium *Bacillus*
 RT *subtilis*.";
 RL Nature 390:249-256(1997).
 RN [9]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;

RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: D78193; BAA11286.1; -;
 DR EMBL: Z99124; CAB16064.1; -;
 KW Complete proteome.
 SQ SEQUENCE 387 AA; 44247 MW; 5EC34B07BA975BB9 CRC64;
 OY 2 SDSFFYTMENS 11
 Db 131 TDGFYTLVNS 140
 RESULT 6
 ID 091TS2 PRELIMINARY; PRT; 636 AA.
 AC 091TS2;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE T27.
 OS Tupala herpesvirus.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae.
 OC NCBI_TaxID=10397;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-2;
 RX MEDLINE=21211637; PubMed=11312357;
 RA Bahr U., Darai G.;
 RT "Analysis and Characterization of the Complete Genome of Tupala (Tree
 RT Shrew) Herpesvirus.";
 RL J. Virol. 75:4854-4870(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-2;
 RA Darai G., Bahr U.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF281817; AAK57065.1; -;
 SQ SEQUENCE 636 AA; 72058 MW; F6489A4BAD99FE85 CRC64;
 OY 1 MSDGFYTMENSDRR 14
 Db 594 MPHGTYIEDDR 607
 Query Match 51.2%; Score 41; DB 12; Length 636;
 Best Local Similarity 57.1%; Pred. No. 69;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 RESULT 7
 ID 09UP56 PRELIMINARY; PRT; 804 AA.
 AC 09UP56;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE KIAA1076 PROTEIN (FRAGMENT).
 GN KIAA1076.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE=99397452; PubMed=10470851;
 RA Kikuno R., Nagase T., Ishikawa K., Hirose M., Miyajima N.,

RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 6:197-205(1999).
 CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
 DR EMBL: AB028999; BAA83028.1; -
 DR InterPro: IPR003616; PostSET.
 DR InterPro: IPR001214; SET.
 DR Pfam: PF00856; SET; 1.
 DR SMART: SM00508; PostSET; 1.
 DR SMART: SM00317; SET; 1.
 DR PROSITE: PS00280; SET; 1.
 FT NON TER
 SO SEQUENCE 804 AA; 87997 MW; 13CAB0BA5420BE67 CRC64;

Query Match 51.2%; Score 41; DB 4; Length 804;
 Best Local Similarity 50.0%; Pred. No. 89;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 2 SDGFTYMSDR 13
 DB 587 SEGFTYIDKDK 598

RESULT 8
 ID 094304 PRELIMINARY; PRT; 822 AA.
 AC 094304;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PUTATIVE S-RECEPTOR KINASE.
 GN P0039A07.12.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Euphorbiaceae; Oryzae; Oryza.
 OX NCB1_TaxID=4530;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. NIPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa niponbare(GAS) genomic DNA, chromosome 1, PAC
 RT clone: P0039A07."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP003235; BAB64106.1; -
 KW Receptor; Kinase.
 SO SEQUENCE 822 AA; 90696 MW; 96C0E1FDACE42F27 CRC64;

Query Match 51.2%; Score 41; DB 10; Length 822;
 Best Local Similarity 77.8%; Pred. No. 91;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 SDGFTYMSDR 10
 DB 360 TDGFTYMAN 368

RESULT 9
 ID 026775 PRELIMINARY; PRT; 2550 AA.
 AC 026775;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE TB-292 MEMBRANE ASSOCIATED PROTEIN.
 OS Trypanosoma brucei.
 OC Eukaryota; Euzoenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCB1_TaxID=5691;
 RN (1)

RP SEQUENCE FROM N.A.
 RC STRAIN-427-60;
 RX MEDLINE-94179225; PubMed-8132566;
 RA Lee M., Russell D., D'Alessandro P., Van der Ploeg L.;
 RT "Identification of membrane associated proteins in Trypanosoma brucei
 RT encoding an internal, EARLRAE amino acid repeat."
 RL J. Biol. Chem. 269:8408-8415(1994).
 DR EMBL: X73956; CAA52142.1; -
 SO SEQUENCE 2550 AA; 289354 MW; 484C297B65A8D376 CRC64;

Query Match 51.2%; Score 41; DB 5; Length 2550;
 Best Local Similarity 46.7%; Pred. No. 3.1e+02;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 MSDGFTYMSDRRK 15
 DB 2237 MADGCVAVSREDRDK 2251

RESULT 10
 ID 054026 PRELIMINARY; PRT; 144 AA.
 AC 054026;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ORF3 PROTEIN.
 GN ORF3.
 OS Paracoccus denitrificans.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 OC Paracoccus.
 OX NCB1_TaxID=266;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-PD1222;
 RA Reijnders W.N.M., Harms N.;
 RT "Paracoccus denitrificans, MxaA, MxaC, MxaK, MxaL, MxaD and orf's."
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ000884; CAA04384.1; -
 FT CHAIN
 SO SEQUENCE 144 AA; 16695 MW; 849D32E80773C780 CRC64;

Query Match 50.0%; Score 40; DB 2; Length 144;
 Best Local Similarity 50.0%; Pred. No. 20;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 2 SDGFTYMSDR 13
 DB 66 TDGYSDNEDR 77

RESULT 11
 ID 09R2J9 PRELIMINARY; PRT; 437 AA.
 AC 09R2J9;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HEMOLYSIN, PUTATIVE.
 GN DRB0128.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCB1_TaxID=129;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-R1;
 RX MEDLINE-20036896; PubMed-10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,

```

RA Vamathevan J.J., Lam P., McDonald L., Uitterbeck T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
DR EMBL: AE001826; AAF12559.1; -.
DR TIGR: DRB0128; -.
DR InterPro: IPR000644; CBS.
DR Pfam: PF00571; CBS; 2.
DR Pfam: PF01595; DUF21.1.
DR SMART: SM00116; CBS; 1.
KW Plasmid; Complete proteome.
SQ SEQUENCE 437 AA; 47593 MW; 1626ABD7440C50A6 CRC64;

Query Match 50.0%; Score 40; DB 16; Length 437;
Best Local Similarity 63.6%; Pred. No. 68;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 SDGFTYMNSD 12
DB 347 SDGFTYMNSD 357

RESULT 12
Q97P13 PRELIMINARY; PRT; 160 AA.
ID 097P13
AC 097P13
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHEICAL PROTEIN SP1831.
GN SP1831.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtaple E., Khouri H., Wolf A.M., Uitterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL: AE007475; AAK75904.1; -.
DR TIGR: SP1831; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 160 AA; 18699 MW; 6ECA8F048E9E4C11 CRC64;

Query Match 48.8%; Score 39; DB 16; Length 160;
Best Local Similarity 46.7%; Pred. No. 34;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 MSDGFTYMNSDRK 15
DB 114 MSDGFTYMNSDRK 128

RESULT 13
Q9HQP0 PRELIMINARY; PRT; 187 AA.

```

```

AC Q9HQP0:
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GLYCEROL-3-PHOSPHATE DEHYDROGENASE CHAIN A.
GN GPDA1 OR VNG1070C.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Shrogha J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Meli R., Goo Y.A.,
RA Lettner B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angelina C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlischer M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL: AE005038; AAG19473.1;
DR InterPro: IPR000447; FAD_Gly3P_dh.
DR PRINTS: PR01001; FADG3PDH.
KW Complete proteome.
SQ SEQUENCE 187 AA; 19735 MW; 933A226B16789311 CRC64;

Query Match 48.8%; Score 39; DB 17; Length 187;
Best Local Similarity 46.2%; Pred. No. 41;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSDGFTYMNSDR 13
DB 103 ISRGFTYVDHADR 115

RESULT 14
Q9VM84 PRELIMINARY; PRT; 306 AA.
ID Q9VM84
AC Q9VM84
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG7328 PROTEIN.
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Gelink R.S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaralides P.G., Scherer S.E., Li P.W., Hostins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Burton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

```


XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 28-APR-1999; 990S-0130891.
PR 30-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 04-MAY-1999; 990S-0132407.
PR 05-MAY-1999; 990S-0132484.
PR 06-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 07-MAY-1999; 990S-0132487.
PR 11-MAY-1999; 990S-0133256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 24-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 29-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.

PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145226.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147933.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148688.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149175.
PR 20-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149928.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.

PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 22-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 23-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 15-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155133.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155655.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160757.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 48.8%; Score 39; DB 21; Length 424;
 Best Local Similarity 50.0%; Pred. NO. 1.5e+02;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Oy 1 MSDGFWTMSD 12
 ||||:|:|:|
 Db 38 msdgysskxtcd 49

RESULT 12
 AAU34516
 ID AAU34516 standard; Protein: 572 AA.

AC AAU34516;

DT 14-FEB-2002 (first entry)

DE E. coli cellular proliferation protein #97.

KW Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

OS Escherichia coli.

PN WO200170955-A2.

XX 27-SEP-2001.

PF 21-MAR-2001; 2001MO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

PA XX

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX WPI: 2001-611495/70.
 DR N-PSDB: AAS52375.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 10109; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 572 AA;

Query Match 48.8%; Score 39; DB 22; Length 572;
 Best Local Similarity 63.6%; Pred. No. 2.1e+02;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 SDGFYTMNSD 12
 1 11:11 1:1
 Db 257 ssqftlmnad 267

RESULT 13
 AAU38403
 ID AU38403 standard; Protein: 572 AA.
 XX
 AC AU38403;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE *Salmonella typhi* cellular proliferation protein #294.
 XX
 KM Antisense: prokaryotic cellular proliferation protein;
 KM antibiotic: antibacterial; drug design.
 XX
 OS *Salmonella typhi*.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.

XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX WPI: 2001-611495/70.
 DR N-PSDB: AAS56262.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 13996; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 572 AA;

Query Match 48.8%; Score 39; DB 22; Length 572;
 Best Local Similarity 63.6%; Pred. No. 2.1e+02;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 SDGFYTMNSD 12
 1 11:11 1:1
 Db 257 ssqftlmnad 267

RESULT 14
 AAG00233
 ID AAG00233 standard; Protein: 100 AA.
 XX
 AC AAG00233;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein, SEQ ID NO: 4314.
 XX
 KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KM gene therapy; chromosome mapping.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 DR WPI: 2000-500381/45.
 DR N-PSDB: AAC00239.

7
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Goddek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kraivit S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., Moleod M.P., McPherson D.,
 RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Stryker R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissendach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,
 RA Gbbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AEO03826; AAF58652.1; -;
 DR FLYBase: FBgn0033627; CG13204.
 DR InterPro: IPR004210; BESS.
 DR Pfam: PF02944; BESS; 1.
 SO SEQUENCE 605 AA; 62998 MW; A44067E859257D12 CRC64;

Query Match 47.8%; Score 43; DB 5; Length 605;
 Best Local Similarity 47.1%; Pred. No. 1.4e+02;
 Matches 8; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 GNAEHVASAVENANRV 17
 Db 366 GLEONLANOVANRLN 382
 RESULT 6
 Q92117 PRELIMINARY; PRT; 676 AA.
 AC Q92117;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE ACYLAMINO-ACID-RELEASING ENZYME.
 GN RC0603.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MALISH 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samson D., Roux V., Cossart P., Weissendach J., Claverie J.-M.,
 RA Reault D.;
 RT Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
 RL Science 293:2093-2098(2001).
 DR EMBL: AEO08621; AAL03141.1; -;
 KW Complete proteome.
 SO SEQUENCE 676 AA; 76356 MW; 9F9977DD060F6246 CRC64;

Query Match 47.8%; Score 43; DB 16; Length 676;

Best Local Similarity 43.8%; Pred. No. 1.6e+02;
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 QY 1 GNAEHVASAVENANRV 16
 Db 639 GRAEHIGDDKNNL1 654

RESULT 7
 Q93129 PRELIMINARY; PRT; 814 AA.
 AC Q93129;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-OCT-2001 (TREMblrel. 18, last annotation update)
 DE NICOTINE DEHYDROGENASE, LARGE CHAIN (EC 1.5.99.4) (NDHC).
 GN NDHC.
 OS Arthropod nictinovorans.
 OC Plasmid PA01.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococcales; Micrococcaceae; Arthropod.
 OX NCBI_TaxID=29320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95115562; PubMed=7815950;
 RA Grether-Beck S., Igloi G., Pust S., Schulz E., Decker K., Brandsch R.;
 RT "Structural analysis and molybdenum-dependent expression of the PA01-
 encoded nicotinic dehydrogenase genes of Arthropod nictinovorans.";
 RL Mol. Microbiol. 13:929-936(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Brandsch R.;
 RL Submitted (Oct-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Brandsch R.;
 RL Submitted (Feb-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: NICOTINE + ACCEPTOR + H(2)O = (S)-6-
 HYDROXYNICOTINE + REDUCED ACCEPTOR.
 CC -1- COFACTOR: MOLYBDOPTERIN.
 CC -1- SUBCELLULAR LOCATION: THE ENZYME IS FOUND IN A SOLUBLE FORM AND IN
 CC A MEMBRANE-ASSOCIATED FORM.
 CC -1- INDUCTION: BY MOLYBDATE.
 CC -1- SIMILARITY: TO THE C-TERMINAL OF VERTEBRATE XANTHINE
 DEHYDROGENASES.
 DR EMBL: X75338; CAAS3088.1; -;
 DR InterPro: IPR000674; Aldxan_dh_C.
 DR Pfam: PF01315; Ald_xan_dh_C; 1.
 DR Pfam: PF02738; Ald_xan_dh_C2; 1.
 KW Oxidoreductase; Molybdenum; Plasmid.
 SO SEQUENCE 814 AA; 87665 MW; 9C5D096EC16028E2 CRC64;

Query Match 47.8%; Score 43; DB 2; Length 814;
 Best Local Similarity 60.0%; Pred. No. 2e+02;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 NAEHVASAVENANRV 16
 Db 184 NAEHVASALYHGRV 198
 RESULT 8
 Q93NH5 PRELIMINARY; PRT; 816 AA.
 AC Q93NH5;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE NDHL.
 OS Arthropod nictinovorans.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Plasmid PA01.

RESULT	12		
09VM45			
09VM45	PRELIMINARY;	PRT;	378 AA.
09VM45			
DT	01-MAY-2000 (TEMBUREl. 13, Created)		
DT	01-MAY-2000 (TEMBUREl. 13, Last sequence update)		
DT	01-MAY-2000 (TEMBUREl. 13, Last annotation update)		
DE	CG8902 PROTEIN.		
GN	CG8902		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		

CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 RN
 RP
 RC SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G.,
 RA Abiri J.F., Agbayani A., An H.-J., Andrews-Plambeck C., Baldwin D.,
 RA Ballev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Doonan K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshfeghi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Styksas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "the genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003616; AAF52482.1; -;
 DR FlyBase: FBgn0031886; CG8902.
 SO SEQUENCE 378 AA; 43809 MW; 0A365CD2B38B2E4 CRC64;

Query Match 46.7%; Score 42; DB 5; Length 378;
 Best Local Similarity 46.7%; Pred. No. 1.2e+02;
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 2 NAEHVASAVENANRV 16
 Db 240 NSOHATALEKATAV 254

RESULT 13
 Q960D9
 ID 0960D9 PRELIMINARY; PRT; 395 AA.
 AC 0960D9;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE SPO5495P.
 GN CG8902.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;

RN
 RP
 RC SEQUENCE FROM N.A.
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunco J., Pacle J., Paragas V., Park S., Phouanavong S., Man K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AV052108; AAK93352.1; -;
 SO SEQUENCE 395 AA; 45690 MW; 402AABID41CE2915 CRC64;

Query Match 46.7%; Score 42; DB 5; Length 395;
 Best Local Similarity 46.7%; Pred. No. 1.2e+02;
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 2 NAEHVASAVENANRV 16
 Db 257 NSOHATALEKATAV 271

RESULT 14
 Q98KR2
 ID 098KR2 PRELIMINARY; PRT; 693 AA.
 AC 098KR2;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 DE NADH-UBIQUINONE DEHYDROGENASE CHAIN 3.
 GN M11362.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxId=381;
 RN
 RP
 RC SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099.
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa K., Kohara M., Matsuno M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP002997; BAB48752.1; -;
 DR InterPro: IPR001041; Ferredoxin.
 DR InterPro: IPR001467; Molybdopterin.
 DR Pfam: PF00111; fer2; 1.
 DR Pfam: PF00384; molybdopterin; 1.
 KW Ubiquinone; Complete proteome.
 SO SEQUENCE 693 AA; 74544 MW; 6CC39A61B38A7631 CRC64;

Query Match 46.7%; Score 42; DB 16; Length 693;
 Best Local Similarity 44.4%; Pred. No. 2.4e+02;
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 1 GNAEHVASAVENANRV 18
 Db 637 GNAEDIAGVAKIGRLNK 654

RESULT 15
 002776
 ID 002776 PRELIMINARY; PRT; 977 AA.
 AC 002776;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE POLY(ADP-RIBOSE) GLYCOPHOLASE.
 GN BPARC.

OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97277328; PubMed=9115250;
RA Lin W., Ame J.C., Aboul-Ela N., Jacobson E.L., Jacobson M.K.;
RT "Isolation and characterization of the cDNA encoding bovine poly(ADP-
RT ribose) glycohydrolase."
RL J. Biol. Chem. 272:11895-11901(1997).
DR EMBL: U78975; AAB53370.1;
KW Hydrolase.
SQ SEQUENCE 977 AA; 110837 MW; 87D2100F979DF377 CRC64;

Query Match 46.7%; Score 42; DB 6; Length 977;
Best Local Similarity 43.8%; Pred. No. 3.5e+02;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
QY 3 AERVASAVENANRVNK 18
:| | : : :| | | :| :
DB 269 SEDVGTGLKNANRLNR 284

Search completed: June 10, 2002, 15:29:18
Job time: 488 sec

1	80	100.0	15	20	AA133650	C. restonii hemoly
2	80	100.0	450	20	AA133651	C. restonii hemoly
3	44	55.0	246	22	ABG11394	Novel human diagnc
4	41	55.0	246	22	ABG29264	Novel human diagnc
5	41	51.2	804	22	ABG03827	Novel human diagnc
6	40	50.0	983	16	AA15751	Novel human diagnc
7	39	48.8	119	22	AAU23768	Novel human enzyme
8	39	48.8	146	22	AAU23244	Novel human enzyme
9	39	48.8	306	22	ABR11549	Novel human enzyme
10	39	48.8	387	21	ABG06340	Drosophila melanog
11	39	48.8	424	21	AA606339	Arabidopsis thalia

PT Protein with hemolytic activity, useful for drugs treating jelly fish
PT strings, pharmaceuticals with blood platelet agglutination activity,
PT pesticides by use of the hemolytic activity, and study of the hemolytic
mechanism -
XX

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 4 GFTYMNSD 12
Db 217 gftymnsd 225

RESULT 4

ABG29264
ID ABG29264 standard; Protein; 246 AA.

AC ABG29264;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #29255.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS93451.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT biodiversity for genetic disorders or other traits and to assess

PS Claim 20; SEQ ID No 59623; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 246 AA;

Query Match 55.0%; Score 44; DB 22; Length 246;
Best Local Similarity 88.9%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 GFTYMNSD 12
Db 217 gftymnsd 225

RESULT 5

ABG03827
ID ABG03827 standard; Protein; 804 AA.

AC ABG03827;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #3818.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS68014.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT biodiversity for genetic disorders or other traits and to assess

PS Claim 20; SEQ ID No 34186; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 804 AA;

Query Match 51.2%; Score 41; DB 22; Length 804;
Best Local Similarity 50.0%; Pred. No. 1,4e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

22-AUG-2000; 2000US-0226661.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 01-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246509.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI: 2001-465566/50.

DR N-PSDB; AAs41638.

Novel polypeptides and polynucleotides useful for diagnosing, preventing, treating neural, immune system, muscular, reproductive, pulmonary, cardiovascular, renal, proliferative disorders and cancerous diseases

Claim 11; SEQ ID NO 1764; 1180pp; English.

PS The present invention relates to the isolation of novel human enzyme
XX polypeptides, and the cDNA (AAs40785-AAs41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 119 AA;

Query Match 48.8%; Score 39; DB 22; Length 119;
Best Local Similarity 63.6%; Pred. No. 38;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 SDGFTYMENS D 12
| | | | | | | |
Db 43 ssgfhtmmnad 53

RESULT 8
AAU23244
ID AAU23244 standard; Protein; 146 AA.
XX
AC AAU23244;
XX
DT 17-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #330.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant;
XX
XX Homo sapiens.
XX
XX WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR -24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.

PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249287.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI: 2001-465566/50.
 DR N-PSDB; AAS41114.
 XX
 PT Novel polypeptides and polynucleotides useful for diagnosing,
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 PT diseases -
 XX
 PS Claim 11: SEQ ID NO 1240; 1180bp; English.
 XX
 CC The present invention relates to the isolation of novel human enzyme
 CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention may comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AAD22915-AAD23814 represent the novel human enzyme polypeptides of the
 CC invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 146 AA:
 QY 2 SDGFTYMNDS 12
 Db 45 ssghfmmnad 55
 Query Match 48.88; Score 39; DB 22; Length 146;
 Best Local Similarity 63.68; Pred. No. 48;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 RESULT 9
 ABB71549
 ID ABB71549 standard; Protein: 306 AA.
 XX
 AC ABB71549;

XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 41439.
 XX
 KM Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PE 23-MAR-2001; 2001MO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR N-PSDB; ABL15652.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 41439; 21bp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL1840-ABL16175) and the encoded proteins
 CC (AB557237-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 306 AA:
 QY 3 DGFTYMNDSRR 14
 Db 21 nsgypmedlrr 32
 Query Match 48.88; Score 39; DB 22; Length 306;
 Best Local Similarity 58.38; Pred. No. 11e+02;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 RESULT 10
 AAG06340
 ID AAG06340 standard; Protein: 387 AA.
 XX
 AC AAG06340;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 3079.
 XX
 KM Protein identification; signal transduction pathway; metabolic pathway;
 KM hybridisation assay; genetic mapping; gene expression control; promoter;
 KM termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 15:29:16 ; Search time 58.44 Seconds

(without alignments)
53.284 Million cell updates/sec

Title: US-09-647-522-2

Sequence: 1 GNAEHVASAVENANRVNK.18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

SPTRMBL.19:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_podent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriaph:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	100.0	450	5	09GV72
2	44	48.9	471	1	004928
3	44	48.9	654	12	089645
4	43	47.8	461	16	0980L5
5	43	47.8	605	5	09V5Y9
6	43	47.8	676	16	092117
7	43	47.8	814	2	059129
8	43	47.8	816	2	093NH5
9	42	46.7	171	2	09F807
10	42	46.7	312	2	B71214
11	42	46.7	312	2	085625
12	42	46.7	378	5	09VM45
13	42	46.7	395	5	0960D9
14	42	46.7	693	16	098KR2
15	42	46.7	977	6	002776
16	41.5	46.1	353	16	09KL63

17	41.5	46.1	461	4	09UT71	09ut71 homo sapien
18	41.5	46.1	915	4	095769	095769 homo sapien
19	41.5	46.1	915	4	09P1S8	09P1S8 homo sapien
20	41.5	46.1	917	4	095937	095937 homo sapien
21	41.5	46.1	917	4	09UT72	09ut72 homo sapien
22	41	45.6	140	15	0905S9	0905S9 human immun
23	41	45.6	411	10	09F5S2	09f5s2 oryza sativ
24	41	45.6	590	16	09J281	09j281 nisseria m
25	41	45.6	590	16	09J0A4	09j0a4 nisseria m
26	41	45.6	631	2	09F4B6	09f4b6 bacillus m
27	41	45.6	631	2	069256	069256 bacillus me
28	41	45.6	693	2	093460	093460 pseudomonas
29	41	45.6	1223	12	09WR70	09wr70 murine hepa
30	41	45.6	1321	12	09J3F1	09j3f1 murine hepa
31	41	45.6	1324	12	039227	039227 murine hepa
32	41	45.6	1324	12	09J3E7	09j3e7 murine hepa
33	41	45.6	1360	12	055253	055253 murine hepa
34	41	45.6	1360	12	09IKD1	09ikd1 rat slaloda
35	41	45.6	1361	12	P90210	P90210 murine hepa
36	41	45.6	1361	12	083331	083331 murine hepa
37	41	45.6	1361	12	066199	066199 human coron
38	41	45.6	1361	12	090D01	090d01 murine hepa
39	41	45.6	1361	12	086623	086623 human coron
40	41	45.6	1363	12	066290	066290 human coron
41	41	45.6	1363	12	066291	066291 human coron
42	41	45.6	1363	12	09QAR5	09qar5 bovine coro
43	41	45.6	1363	12	090A08	090a08 bovine coro
44	41	45.6	1363	12	091A26	091a26 bovine coro
45	41	45.6	1376	12	083330	083330 murine hepa

ALIGNMENTS

RESULT 1	09GV72	PRELIMINARY;	PRT;	450 AA.
ID	09GV72	AC	09GV72;	
AC	09GV72	01-MAR-2001 (TREMBlrel. 16, Created)		
DT	01-MAR-2001 (TREMBlrel. 16, Last sequence update)			
DT	01-OC-2001 (TREMBlrel. 18, Last annotation update)			
DE	TOXIN-1.			
GN	CRT-1.			
OS	Carybdea rastoni.			
OC	Eukaryota; Metazoa; Cnidaria; Cubozoa; Cubomedusae; Carybdeidae;			
OC	Carybdea.			
OX	NCBI_TaxID=78582;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20422301; PubMed=10964707;			
RA	Nagai H., Takawa K., Nakao M., Ito E., Miyake M., Noda M., Nakajima T.,			
RT	"Novel Proteaceous toxins from the box jellyfish (sea wasp) Carybdea rastoni."			
RT	Biochem. Biophys. Res. Commun. 275:582-588(2000).			
RL	EMBL: AB015878; BAB12728.1;			
DR	EMBL: AB015878; BAB12728.1;			
SQ	SEQUENCE 450 AA; 49392 MW; CD393CF25BEFDZFD CRC64;			

Query Match 100.0%; Score 90; DB 5; Length 450;
Best Local Similarity 100.0%; Pred. No. 2.3e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GNAEHVASAVENANRVNK 18
DB	250	GNAEHVASAVENANRVNK 267
RESULT 2		
ID	004928	PRELIMINARY;
AC	004928;	PRT; 471 AA.
DT	01-NOV-1996 (TREMBlrel. 01, Created)	

DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE (CSM3) PLASMID-RELATED CHROMOSOMAL DNA SEQUENCE CFR-II.
 OS Methanothermobacter thermautotrophicus.
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
 OC Methanothermobacter.
 OX NCBI_TaxID=145262;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=93341463; PubMed=8393514;
 RX Noelling J., Van Eden F.J.M., de Vos W.M.;
 RT "Distribution and characterization of plasmid-related sequence in the
 RT chromosomal DNA of thermophilic Methanobacterium strain".
 RL Mol. Gen. Genet. 240:81-91(1993).
 DR EMBL; X69114; CAA4867.1; -
 DR InterPro: IPR001434; DUF11.
 DR Pfam: PF01345; DUF11. 2.
 SQ SEQUENCE 471 AA; 51595 MW; 44B8DB34FAB57BDE CRC64;

Query Match 48.9%; Score 44; DB 1; Length 471;
 Best Local Similarity 56.2%; Pred. No. 72;
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 OY 2 GNAEHVASAVENANRVN 17
 Db 325 NASAVSAVDNSNVN 340

RESULT 3
 ID 089645 PRELIMINARY; PRT; 654 AA.
 AC 089645;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE ORF2.
 OS Barley yellow dwarf virus (isolate NY-RPV) (BYDV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
 OC Luteovirus.
 OX NCBI_TaxID=12039;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91108372; PubMed=2273382;
 RX Vincent J.R., Deng P.P., Lister R.M., Larkins B.A.;
 RT "Nucleotide sequences of coat protein genes for three isolates of
 RT barley yellow dwarf virus and their relationships to other luteoviruses
 RT coat protein sequences".
 RL J. Gen. Virol. 71:2791-2799(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92013950; PubMed=1840612;
 RX Larkins B.A., Lister R.M., Vincent J.R.;
 RT "Nucleotide sequence analysis and genomic organization of the NY-RPV
 RT isolate of barley yellow dwarf virus".
 RL J. Gen. Virol. 72:2347-2355(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NY;
 RA Beckett R.J., Vincent J.R., Lister R.M., Miller W.A.;
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
 DR EMBL; D10206; BA01053.1; -
 DR MEROPS; S52.001; -
 DR InterPro: IPR000382; Luteo_ORF2.
 DR Pfam: PF02122; Luteo_ORF2. 1.
 DR PRINTS: PRO0913; LVIRSORF2.
 SQ SEQUENCE 654 AA; 70566 MW; 53D5E868926D3D28 CRC64;

Query Match 48.9%; Score 44; DB 12; Length 654;
 Best Local Similarity 55.6%; Pred. No. 1e+02;
 Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 1 GNAEHVASAVENANRVN 18
 Db 458 GNGKAASAVTNAAANK 475

RESULT 4
 ID 0980L5 PRELIMINARY; PRT; 461 AA.
 AC 0980L5;
 DT 01-OCT-2001 (TREMblrel. 18, Created)
 DT 01-OCT-2001 (TREMblrel. 18, last sequence update)
 DT 01-OCT-2001 (TREMblrel. 18, last annotation update)
 DE ABC TRANSPORTER XYLOSE-BINDING LIPOPROTEIN.
 GN MYPQ_3460.
 OS Mycoplasma pulmonis.
 OC Bacteria; Firmicutes; Bacillus/clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2107;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UAB CTIP;
 RX MEDLINE=21267165; PubMed=11353084;
 RA Chandaud I., Helling R., Ferris S., Barbe V., Samson D., Galisson F.,
 RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
 RA Blanchard A.;
 RT "The complete genome sequence of the murine respiratory pathogen
 RT Mycoplasma pulmonis".
 RL Nucleic Acids Res. 29:2145-2153(2001).
 DR EMBL; AL445564; CAC13519.1; -
 DR MYPULIST; MYPUL_3460; -
 DR InterPro: IPR003760; Bmp.
 DR Pfam: PF02608; Bmp; 1.
 DR KW Complete proteome.
 SQ SEQUENCE 461 AA; 51096 MW; 6724D4D820809CE4 CRC64;

Query Match 47.8%; Score 43; DB 16; Length 461;
 Best Local Similarity 53.3%; Pred. No. 1e+02;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 OY 1 GNAEHVASAVENANRVN 15
 Db 441 GDLEHLKSTVOKANR 455

RESULT 5
 ID 09V5Y9 PRELIMINARY; PRT; 605 AA.
 AC 09V5Y9;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE CG13204 PROTEIN.
 GN CG13204.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazer J.G., Champé M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Adwayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 15:29:41 ; Search time 18.65 Seconds
(without alignments)
37.370 Million cell updates/sec

Title: US-09-647-522-2

Perfect score: 90
Sequence: 1 GNAEHVASAVENANRVNK 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	47.8	142	1 VG09_BPPI	P40820 bacteriophage
2	43	47.8	194	1 AX2A_PHAU	P32293 phaseolus a
3	43	47.8	249	1 TPIS_EMENI	P04828 emericella
4	43	47.8	397	1 PACA_BPPI	P28813 bacteriophage
5	42	46.7	163	1 YAJ0_HAEIN	P44096 haemophilus
6	42	46.7	481	1 ABGB_ECOLI	P76052 escherichia
7	41.5	46.1	915	1 SMOO_HUMAN	P53814 homo sapien
8	41	45.6	1235	1 VGL2_CVMB	P11225 murine coro
9	41	45.6	1324	1 VGL2_CVMB	P11225 murine coro
10	41	45.6	1353	1 VGL2_CVMB	P36334 human coro
11	41	45.6	1363	1 VGL2_CVMB	P25190 bovine coro
12	41	45.6	1363	1 VGL2_CVMB	P25191 bovine coro
13	41	45.6	1363	1 VGL2_CVMB	P25192 bovine coro
14	41	45.6	1363	1 VGL2_CVMB	P15777 bovine coro
15	41	45.6	1363	1 VGL2_CVMB	P25193 bovine coro
16	41	45.6	1363	1 VGL2_CVMB	P25194 bovine coro
17	41	45.6	1376	1 VGL2_CVMB	P22432 murine coro
18	41	45.6	1376	1 VGL2_CVMB	P02385 murine coro
19	40	44.4	231	1 KC65_ARCFU	O29003 archaeoglob
20	40	44.4	382	1 FORA_MERTH	P56810 methanobact
21	40	44.4	465	1 MTR6_HUMAN	O99217 homo sapien
22	40	44.4	538	1 THSB_MERTH	O26885 methanobact
23	40	44.4	631	1 MERA_BACSR	P16171 bacillus sp
24	40	44.4	1159	1 Y196_HUMAN	O12768 homo sapien
25	39	43.3	297	1 STX4_HUMAN	O12846 homo sapien
26	39	43.3	418	1 FTS2_TREPA	O83405 treponema p
27	39	43.3	477	1 DLDH_HAEIN	P43784 haemophilus
28	39	43.3	582	1 ASN2_PEA	P19292 pisum sativ
29	39	43.3	733	1 MK10_YEAST	O02137 saccharomyc
30	39	43.3	936	1 RPOB_CLAPU	P22372 clauvaceps p
31	38	42.2	230	1 YJ02_MERTH	O27924 methanobact
32	38	42.2	258	1 I435_SOLTU	P93784 solanum tub
33	38	42.2	258	1 I436_LYVCS	P93211 lycopersico

ALIGNMENTS

RESULT	ID	Score	Match	Length	ID	Description
1	VG09_BPPI	47.8%	142	142	1	SOHB_VIBCH
2	AX2A_PHAU	69.2%	194	194	1	RS2_COXBU
3	TPIS_EMENI	69.2%	249	249	1	TRPD_AERPE
4	PACA_BPPI	69.2%	397	397	1	Y148_MYCCE
5	YAJ0_HAEIN	69.2%	163	163	1	AAFC_YEAST
6	ABGB_ECOLI	69.2%	481	481	1	MESJ_BUCAT
7	SMOO_HUMAN	69.2%	915	915	1	TRME_HELPJ
8	VGL2_CVMB	69.2%	1235	1235	1	PEP_DROME
9	VGL2_CVMB	69.2%	1324	1324	1	HOX1_HAIRO
10	VGL2_CVMB	69.2%	1353	1353	1	RA50_THEMA
11	VGL2_CVMB	69.2%	1363	1363	1	ORC1_HUMAN
12	VGL2_CVMB	69.2%	1363	1363	1	ORC1_HUMAN
13	VGL2_CVMB	69.2%	1363	1363	1	ORC1_HUMAN
14	VGL2_CVMB	69.2%	1363	1363	1	ORC1_HUMAN
15	VGL2_CVMB	69.2%	1363	1363	1	ORC1_HUMAN
16	VGL2_CVMB	69.2%	1363	1363	1	ORC1_HUMAN
17	VGL2_CVMB	69.2%	1376	1376	1	ORC1_HUMAN
18	VGL2_CVMB	69.2%	1376	1376	1	ORC1_HUMAN
19	KC65_ARCFU	69.2%	231	231	1	ORC1_HUMAN
20	FORA_MERTH	69.2%	382	382	1	ORC1_HUMAN
21	MTR6_HUMAN	69.2%	465	465	1	ORC1_HUMAN
22	THSB_MERTH	69.2%	538	538	1	ORC1_HUMAN
23	MERA_BACSR	69.2%	631	631	1	ORC1_HUMAN
24	Y196_HUMAN	69.2%	1159	1159	1	ORC1_HUMAN
25	STX4_HUMAN	69.2%	297	297	1	ORC1_HUMAN
26	FTS2_TREPA	69.2%	418	418	1	ORC1_HUMAN
27	DLDH_HAEIN	69.2%	477	477	1	ORC1_HUMAN
28	ASN2_PEA	69.2%	582	582	1	ORC1_HUMAN
29	MK10_YEAST	69.2%	733	733	1	ORC1_HUMAN
30	RPOB_CLAPU	69.2%	936	936	1	ORC1_HUMAN
31	YJ02_MERTH	69.2%	230	230	1	ORC1_HUMAN
32	I435_SOLTU	69.2%	258	258	1	ORC1_HUMAN
33	I436_LYVCS	69.2%	258	258	1	ORC1_HUMAN

Query Match: 47.8%; Score 43; DB 1; Length 142;

Best Local Similarity: 69.2%; Pred. No. 3.4;

Matches: 9; Conservative: 2; Mismatches: 2; Indels: 0; Gaps: 0;

RESULT 2

AX2A_PHAU STANDARD; PRT: 194 AA.

AC P32293; 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

GN AUX22A OR ARG3.

OS Phaseolus aureus (Mung bean) (Vigna radiata).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Query Match	47.8%	Score 43	DB 1	Length 19
Best Local Similarity	52.9%	Pred. No. 4	8	
Matches	9	Conservative	3	Mismatches 5
				Indels 0
				Gaps 0
QY	2	NAEHVAVENANRVNK	18	
Db	19	DAEHVAVANKGEKKNK	35	

	RESULT	3
ID	TPIS_EMENT	
AC	TPIS_EMENT	STANDARD:
P04828;		PRT; 249 AA.
DT	13-AUG-1987	(Rel. 05, Created)
DT	13-AUG-1987	(Rel. 05, Last sequence update)
DT	01-FEB-1996	(Rel. 33, Last annotation update)
DE	Triosephosphate isomerase	(EC 5.3.1.1) (TIM).
GN	TPIA.	
OS	Emeritella nidulans (Aspergillus nidulans).	
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;	
OC	Eurotiales; Trichocomaceae; Emeritella.	
OX	NCBI_TaxID=5072;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=66245050; PubMed=3521890;	
RA	McKnight G.L., O'Hara P.J., Parker M.L.;	
RT	"Nucleotide sequence of the triosephosphate isomerase gene from	
RT	Aspergillus nidulans: implications for a differential loss of	
RT	introns.";	
RL	Cell 46:143-147(1986).	
CC	-I CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate -> glycero-	
CC	phosphate.	
CC	-I PATHWAY: PLAYS AN IMPORTANT ROLE IN SEVERAL METABOLIC PATHWAYS.	
CC	-I SUBUNIT: HOMODIMER.	
CC	-I SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	

DR	EMBL	D10019	BAA00908.1	-
DR	PIR	A25502	ISASTN	
DR	HSSP	P00940	RTIM	
DR	InterPro	IPR000652	Trioseph_Isomerase	
DR	Pfam	PF00121	TIM	1
DR	ProDom	PD001005	Trioseph_Isomrse	1
DR	PROSITE	PS00171	TIM	1
KW	Isomerase	Glycolysis	Gluconeogenesis	Fatty acid biosynthesis
KW				Penicose shunt
FT	ACT_SITE	94		BY SIMILARITY
FT	ACT_SITE	166		BY SIMILARITY
SQ	SEQUENCE	249 AA	27157 MW	03AFLBA27B545A1C CRC64

```

Query Match      47.8%  Score 43;  DB 1;  Length 249;
Best Local Similarity 47.1%  Pred. No. 6,2;
Matches      8;  Conservative      2;  Mismatches      7;  Indels      0;  Gaps      0;

Oy'      1  GNAEHVAVSENNANRVN 17
          ||||  |::|  |
          15  GNAESTTIIKNNLSAN 31

```

	RESULT	4	
PACB_BPPI	ID	PACB_BPPI	STANDARD;
AC	P28813;		PRT; 397 AA.
DT	01-DEC-1992 (Rel. 24, Created)		
DT	01-DEC-1992 (Rel. 24, Last sequence update)		
DT	01-DEC-1992 (Rel. 24, Last annotation update)		
DE	Teminase A protein (PACASE A protein) (DNA packaging A protein).		
GN	PACB.		
OS	Bacteriophage P1.		
OC	Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae.		
OX	NCBI_Taxid=10678;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=92167273; PubMed=1538406;		
RA	Skorupski K., Pierce J.C., Sauer B., Sternberg N.;		
RT	"bacteriophage P1 genes involved in the recognition and cleavage of		
RL	the phage packaging site (pac).";		
J. Mol. Biol.	223:977-989(1992)."		
-I-	FUNCTION: NECESSARY FOR RECOGNITION AND CLEAVAGE OF THE PHAGE		
PACKAGING SITE (PAC), TOGETHER WITH THE PACB PROTEIN. BOTH ARE			
PROBABLY THE TWO MAJOR SUBUNITS OF THE PHAGE PACASE.			
-I- SUBUNIT: MULTIMER OF TWO DISTINCT SUBUNITS (BT SIMILARITY).			
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; M74046; AAA21723.1; -		
DR	PIR; S21780; S21780.		
KM	Hydrolase; Nuclease; DNA-binding.		
QO	SEQUENCE 397 AA; 45279 MW; 247B64DE3E2D550 CRC64;		

Query Match	47.8%	Score 43	DB 1	Length 397
Best Local Similarity	69.2%	Pred. No. 10		
Matches	9	Conservative	2	Mismatches 0
				Indels 0
				Gaps 0
OY	6	VASAVENANRVNK	18	
		:::		
Ob	93	VSAAVENAKRQNK	105	

```

RESULT 5
ID YAJQ_HAEIN STANDARD; PRT; 163 AA.
AC P44096;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein H11034.
GN H11034.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Uettermann T.R., Hanna M.C., Spriggs T., Hedblom E., Cotton M.D.,
RA Elze L.D., Fritchman J.L., Fuhrmann J.T., Saudek D.M., Brandon R.C.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd.";
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429(2000).
CC -1- SIMILARITY: STRONG. TO E.COLI YAJQ.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U33784; AAC22694.1; -
DR TIGR: H11034; -
KW Complete proteome.
SQ SEQUENCE 163 AA; 18551 MW; 1831F75A87A74717 CRC64;

Query Match 46.7%; Score 42; DB 1; Length 163;
Best Local Similarity 81.8%; Pred. No. 5.8;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 6 VASAVENANRV 16
Db 15 VRNAVENANRV 25

```

```

RESULT 6
ID ABGB_ECOLI STANDARD; PRT; 481 AA.
AC P76052;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aminobenzoyl-glutamate utilization protein B.
GN ABGB OR B1337.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K-12;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP CHARACTERIZATION.
RC STRAIN-BN101;
RX Hussein M.J., Green J.M., Nichols B.P.;
RA "Characterization of mutations that allow p-aminobenzoyl-glutamate
utilization by Escherichia coli.";
RL J. Bacteriol. 180:6260-6268(1998).
CC -1- FUNCTION: REQUIRED BUT NOT ESSENTIAL FOR AMINO BENZOYL-GLUTAMATE
UTILIZATION. MAY PARTICIPATE IN HYDROLYSIS OF AMINO BENZOYL-
GLUTAMATE TO AMINO BENZOATE, EITHER ALONE OR IN COMBINATION WITH
ABGB.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000231; AAC74419.1; -
DR Ecogene; EC13351; abgb.
KW Complete proteome.
SQ SEQUENCE 481 AA; 52193 MW; 84B09100A563E07C CRC64;

Query Match 46.7%; Score 42; DB 1; Length 481;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 NAEHVASAVENA 13
Db 39 SAEHLASALESA 50

```

```

RESULT 7
ID SMOO_HUMAN STANDARD; PRT; 915 AA.
AC P53814; 000569;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Smoochelin.
GN SMTN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG FORM).
RC TISSUE-Vascular smooth muscle;
RA van Eys G.J.J.M.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE-Smooth muscle;
RX MEDLINE=96295554; PubMed=8707825;
RA van der Loop F.T.L., Schaart G., Timmer E.D.J., Ramaekers F.C.S.,
RA van Eys G.J.J.M.;
RT "Smoothenin, a novel cytoskeletal protein specific for smooth muscle

```

RT cells.",
CC J. Cell Biol. 134:401-411(1996).
CC -1- FUNCTION: STRUCTURAL PROTEIN (CYTOSKELETAL).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; EXHIBITS A FILAMENTOUS
CC ORGANIZATION.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS: A LONG FORM (SHOWN
CC HERE) AND A SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE; CONTRACTILE OR VASCULAR (FOR
CC THE LONG FORM).
CC -1- SIMILARITY: CONTAINS 1 CALPONTIN-HOMOLOGY (CH) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y13492; CAA73884.1; -
DR EMBL: Z49989; CAA90281.1; -
DR HSBP: Q01082; IBKR.
DR MIM: 602127; -
DR InterPro: IPR001715; Calponin_hom.
DR Pfam: PF00307; CH.1.
DR SMART: SM00033; CH.1.
DR PROSITE: PS50021; CH.1
DR Structural protein; Alternative splicing.
FT DOMAIN 708 713 POLY-SER.
FT DOMAIN 778 809 PRO-RICH.
FT CARBOHYD 795 901 CH.
FT VARPPLIC 1 544 MISSING (IN SHORT ISOFORM).
SO SEQUENCE 915 AA; 98449 MW; 338558B49E371B8 CRC64.

Query Match 46.1%; Score 41.5; DB 1; Length 915;
Best Local Similarity 57.1%; Pred. No. 44;
Matches 12; Conservative 2; Mismatches 2; Indels 5; Gaps 1;

QY 3 AEHVASAVENAN-----RYNK 18
DB 550 AEPLAAVEANGAEARARVKN 570

RESULT 8
VGL2_CVMJH STANDARD; PRT; 1235 AA.
AC P11225;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE E2 glycoprotein precursor (Spike glycoprotein) (Peplomer protein)
DE [Contains: Spike protein S1 (90B); Spike protein S2 (90A)].
GN S.
OS Murine coronavirus MHV (strain JHM).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_Taxid=11144;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8711467; PubMed=3027248;
RA Schmidt I., Skinner M.A., Siddell S.G.;
RT "Nucleotide sequence of the gene encoding the surface projection
RT glycoprotein of coronavirus MHV-JHM.";
RT J. Gen. Virol. 68:47-56(1987).
RL -1- FUNCTION: THE PEPLMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION
CC AND IN SYNCYTIIUM FORMATION.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04797; CAA28484.1; -
DR EMBL: D00093; BAA00057.1; -
DR PIR: A33095; YGTHM7.
DR InterPro: IPR002552; Corona_S2.
DR Pfam: PF01601; Corona_S2; 2.
DR Glycoprotein; Envelope protein; Transmembrane; signal.
FT SIGNAL 1 10
FT CHAIN 11 1235
FT CHAIN 11 628
FT CHAIN 629 1235
FT DOMAIN 11 1174
FT TRANSMEM 1175 1197
FT DOMAIN 1198 1235
FT DOMAIN 1198 1215
FT CARBOHYD 31 31
FT CARBOHYD 60 60
FT CARBOHYD 134 134
FT CARBOHYD 192 192
FT CARBOHYD 357 357
FT CARBOHYD 435 435
FT CARBOHYD 536 536
FT CARBOHYD 568 568
FT CARBOHYD 576 576
FT CARBOHYD 599 599
FT CARBOHYD 648 648
FT CARBOHYD 665 665
FT CARBOHYD 804 804
FT CARBOHYD 1091 1091
FT CARBOHYD 1101 1101
FT CARBOHYD 1120 1120
FT CARBOHYD 1136 1136
FT CARBOHYD 1157 1157
SO SEQUENCE 1235 AA; 136653 MW; 25962AD6C1F92DD2 CRC64.

Query Match 45.6%; Score 41; DB 1; Length 1235;
Best Local Similarity 53.8%; Pred. No. 73;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 GNAHVASAVENA 13
DB 1005 GNGNHITSLVONA 1017

RESULT 9
VGL2_CVMAS STANDARD; PRT; 1324 AA.
AC P11224;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE E2 glycoprotein precursor (Spike glycoprotein) (Peplomer protein)
DE [Contains: Spike protein S1 (90B); Spike protein S2 (90A)].
GN S.
OS Murine coronavirus MHV (strain A59).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_Taxid=11142;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88072088; PubMed=2825419;
RA van der Zeijst B.A.M., Horzinek M.C., Spaan W.J.M.;
RT "Primary structure of the glycoprotein E2 of coronavirus MHV-A59 and
RT identification of the trypsin cleavage site.";
RT Virology 161:479-487(1987).
RL -1- FUNCTION: THE PEPLMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

RT   coronavirus OC43.3;
RL   J. Gen. Virol. 74:1981-1987(1993).
CC   -1- FUNCTION: THE PEPLIMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC   TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION
CC   AND IN SYNCYTIAL FORMATION.
CC   -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL: L14643; AAA03055.1; -.
DR   PIR: J02168; J02168.
DR   InterPro: IPR002352; Corona_S2.
KW   Glycoprotein; Envelope protein; Transmembrane; Signal.
FT   SIGNAL          1      17
FT   CHAIN           18     1353
FT   CHAIN           18     757
FT   CHAIN           758     1353
FT   DOMAIN           18     1298
FT   TRANSMEM       1299     1318
FT   DOMAIN           1319     1353
FT   DOMAIN           1319     1336
FT   CARBOHYD        59
FT   CARBOHYD        133     133
FT   CARBOHYD        146     146
FT   CARBOHYD        202     202
FT   CARBOHYD        363     363
FT   CARBOHYD        441     441
FT   CARBOHYD        496     496
FT   CARBOHYD        639     639
FT   CARBOHYD        666     666
FT   CARBOHYD        704     704
FT   CARBOHYD        729     729
FT   CARBOHYD        778     778
FT   CARBOHYD        927     927
FT   CARBOHYD        1184     1184
FT   CARBOHYD        1214     1214
FT   CARBOHYD        1224     1224
FT   CARBOHYD        1243     1243
FT   CARBOHYD        1257     1257
FT   CARBOHYD        1278     1278
SO   SEQUENCE        1353 AA; 150088 MW; 7BICE3CBDB1A902A CRC64;

Query Match      45.6%; Score 41; DB 1; Length 1353;
Best Local Similarity 53.8%; Pred. No. 80;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy      1  GNAEHVASAVENA 13
Db      1128  GNGNHITSLVONA 1140

RESULT 11
VGL2_CVBF
VGL2_CVBF      STANDARD;      PRT; 1363 AA.
AC      P25190;
DT      01-MAY-1992 (Rel. 22, Created)
DT      01-MAY-1992 (Rel. 22, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      E2 glycoprotein precursor (Spike glycoprotein) (Peplomer protein)
DE      [contains: Spike protein S1 (90B); Spike protein S2 (90A)].
CN      S.
CS      Bovine coronavirus (strain F15).
CC      Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
CC      Coronaviridae; Coronavirus.

```

```

OX NCBI_TaxID-11129;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-90171938; PubMed-2155300;
RA Boiteau P., Cruciere C., Laporte J.;
RT "Nucleotide sequence of the glycoprotein S gene of bovine enteric
RT coronavirus and comparison with the S proteins of two mouse hepatitis
RT virus strains.";
RL J. Gen. Virol. 71:487-492(1990).
CC -I- FUNCTION: THE PEPLIMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, D00731; BAA00631.1; -
DR PIR, A34151; VGIHFL.
DR InterPro: IPR002552; Corona_S2.
DR Pfam, PF01601; Corona_S2. 2.
KW Glycoprotein; Envelope protein; Transmembrane; Signal.
FT SIGNAL 1 17
FT CHAIN 18 1363
FT CHAIN 18 768
FT CHAIN 769 1363
FT DOMAIN 18 1311
FT TRANSMEM 1312 1328
FT DOMAIN 1329 1363
FT DOMAIN 1329 1346
FT CARBOHYD 59 59
FT CARBOHYD 133 133
FT CARBOHYD 198 198
FT CARBOHYD 359 359
FT CARBOHYD 437 437
FT CARBOHYD 649 649
FT CARBOHYD 676 676
FT CARBOHYD 696 696
FT CARBOHYD 714 714
FT CARBOHYD 739 739
FT CARBOHYD 788 788
FT CARBOHYD 937 937
FT CARBOHYD 1194 1194
FT CARBOHYD 1224 1224
FT CARBOHYD 1234 1234
FT CARBOHYD 1253 1253
FT CARBOHYD 1267 1267
FT CARBOHYD 1288 1288
SQ SEQUENCE 1363 AA; 150746 MW; 3245CD74D70AD135 CRC64;

Query Match 45.6%; Score 41; DB 1; Length 1363;
Best Local Similarity 53.8%; Pred. No. 81;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 GNAEHVASAVENA 13
DB 1138 GNGNHISLYONA 1150

RESULT 12
VGL2_CVBL9
ID VGL2_CVBL9 STANDARD; PRT; 1363 AA.
AC P25191;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE E2 glycoprotein precursor (Spike glycoprotein) (Peplomer protein)
DE [Contains: Spike protein S1 (90B); Spike protein S2 (90A)].
GN S.

```

```

OS Bovine coronavirus (strain L9).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID-11130;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91272503; PubMed-2053289;
RA Zhang X., Kousoulas K.G., Storz J.;
RT "Comparison of the nucleotide and deduced amino acid sequences of the
RT S genes specified by virulent and avirulent strains of bovine
RT coronaviruses.";
RL Virology 183:397-404(1991).
CC -I- FUNCTION: THE PEPLIMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M4667; AAA42907.1; -
DR PIR, A40320; VGIH9.
DR InterPro: IPR002552; Corona_S2.
DR Pfam, PF01601; Corona_S2. 2.
KW Glycoprotein; Envelope protein; Transmembrane; Signal.
FT SIGNAL 1 17
FT CHAIN 18 1363
FT CHAIN 18 768
FT CHAIN 769 1363
FT DOMAIN 18 1311
FT TRANSMEM 1312 1328
FT DOMAIN 1329 1363
FT CARBOHYD 59 59
FT CARBOHYD 133 133
FT CARBOHYD 198 198
FT CARBOHYD 359 359
FT CARBOHYD 437 437
FT CARBOHYD 649 649
FT CARBOHYD 676 676
FT CARBOHYD 696 696
FT CARBOHYD 714 714
FT CARBOHYD 739 739
FT CARBOHYD 788 788
FT CARBOHYD 937 937
FT CARBOHYD 1194 1194
FT CARBOHYD 1224 1224
FT CARBOHYD 1234 1234
FT CARBOHYD 1253 1253
FT CARBOHYD 1267 1267
FT CARBOHYD 1288 1288
SQ SEQUENCE 1363 AA; 150806 MW; 5D27D1A58BFC951 CRC64;

Query Match 45.6%; Score 41; DB 1; Length 1363;
Best Local Similarity 53.8%; Pred. No. 81;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 GNAEHVASAVENA 13
DB 1138 GNGNHISLYONA 1150

RESULT 13
VGL2_CVBLX
ID VGL2_CVBLX STANDARD; PRT; 1363 AA.
AC P25192; Q9GAS2;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE E2 glycoprotein precursor (Spike glycoprotein) (Peplomer protein)
GN S.

```

[Contains: Spike protein S1 (90B); Spike protein S2 (90A)].

DE GN S.

OS Bovine coronavirus (strain LR-138).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

OC Coronaviridae; Coronavirus.

OX NCBI_TaxID=11131;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=91272503; PubMed=2053289;

RA Zhang X., Kousoulas K.G., Storz J.;

RT "Comparison of the nucleotide and deduced amino acid sequences of the

RT S genes specified by virulent and avirulent strains of bovine

RT coronavirus.";

RL Virology 183:397-404(1991).

CC -1- FUNCTION: THE PEPLIMER PROTEIN MEDIATES THE BINDING OF VIRIONS

CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

DR EMBL: AF058942; AAF25499.1; -

DR PIR: C40320; VGIRLY.

DR InterPro: IPR002552; Corona_S2.

DR Pfam: PF01601; Corona_S2; 2.

KW Glycoprotein; Envelope protein; Transmembrane; Signal.

FT SIGNAL 1 17

FT CHAIN 18 1363

FT CHAIN 18 768

FT CHAIN 769 1363

FT DOMAIN 18 1311

FT TRANSEM 1312 1328

FT DOMAIN 1329 1363

FT DOMAIN 1329 1346

FT CARBOHYD 59 59

FT CARBOHYD 133 133

FT CARBOHYD 198 198

FT CARBOHYD 359 359

FT CARBOHYD 437 437

FT CARBOHYD 649 649

FT CARBOHYD 676 676

FT CARBOHYD 696 696

FT CARBOHYD 714 714

FT CARBOHYD 739 739

FT CARBOHYD 788 788

FT CARBOHYD 937 937

FT CARBOHYD 1194 1194

FT CARBOHYD 1224 1224

FT CARBOHYD 1234 1234

FT CARBOHYD 1253 1253

FT CARBOHYD 1267 1267

FT CARBOHYD 1288 1288

SO SEQUENCE 1363 AA; 150676 MW; CE165F733EBF6F7 CRC64;

Query Match 45.6%; Score 41; DB 1; Length 1363;

Best Local Similarity 53.8%; Pred. No. 81;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

RESULT 14

VG2_CVBM STANDARD; PRT; 1363 AA.

AC P15777;

DT 01-APR-1990 (Rel. 14, Created)

01-APR-1990 (Rel. 14, Last sequence update)

16-OCT-2001 (Rel. 40, Last annotation update)

E2 glycoprotein precursor (Spike glycoprotein) (peplomer protein)

[Contains: Spike protein S1 (90B); Spike protein S2 (90A)].

S.

OS Bovine coronavirus (strain Mebus).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

OC Coronaviridae; Coronavirus.

OX NCBI_TaxID=11132;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=90323743; PubMed=2184576;

RA Abraham S., Kienzle T.E., Lapps W.E., Brian D.A.;

RT "Deduced sequence of the bovine coronavirus spike protein and

RT identification of the internal proteolytic cleavage site.";

RL Virology 176:296-301(1990).

CC [2]

CC SEQUENCE OF 1254-1363 FROM N.A.

CC MEDLINE=90320120; PubMed=2142556;

CC Abraham S., Kienzle T.E., Lapps W.E., Brian D.A.;

CC "Sequence and expression analysis of potential nonstructural proteins

CC of 4.9, 4.8, 12.7, and 9.5 kDa encoded between the spike and membrane

CC protein genes of the bovine coronavirus.";

CC Virology 177:488-495(1990).

CC -1- FUNCTION: THE PEPLIMER PROTEIN MEDIATES THE BINDING OF VIRIONS

CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

DR EMBL: M31053; AAA6399.1; -

DR EMBL: M31054; AAA42910.1; -

DR PIR: A34607; VGIRLM.

DR InterPro: IPR002552; Corona_S2.

DR Pfam: PF01601; Corona_S2; 2.

KW Glycoprotein; Envelope protein; Transmembrane; Signal.

FT SIGNAL 1 17

FT CHAIN 18 1363

FT CHAIN 18 768

FT CHAIN 769 1363

FT DOMAIN 18 1311

FT TRANSEM 1312 1328

FT DOMAIN 1329 1363

FT DOMAIN 1329 1346

FT CARBOHYD 59 59

FT CARBOHYD 133 133

FT CARBOHYD 198 198

FT CARBOHYD 359 359

FT CARBOHYD 437 437

FT CARBOHYD 649 649

FT CARBOHYD 676 676

FT CARBOHYD 696 696

FT CARBOHYD 714 714

FT CARBOHYD 739 739

FT CARBOHYD 788 788

FT CARBOHYD 937 937

FT CARBOHYD 1194 1194

FT CARBOHYD 1224 1224

FT CARBOHYD 1234 1234

FT CARBOHYD 1253 1253

FT CARBOHYD 1267 1267

FT CARBOHYD 1288 1288

SO SEQUENCE 1363 AA; 150810 MW; 864B3AE08232F458 CRC64;

Query Match 45.6%; Score 41; DB 1; Length 1363;

Best Local Similarity 53.8%; Pred. No. 81;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 1 GNAEHVASAVENA 13
Db 1138 GNGNHITSLVQNA 1150

RESULT 15

VG12_CVBO STANDARD; PRT; 1363 AA.

AC P25193;

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE E2 glycoprotein precursor (Spike glycoprotein) (Peplomer protein)

DE [Contains: Spike protein S1 (908); Spike protein S2 (90A)].

GN S.

OS Bovine coronavirus (strain Quebec).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

OC Coronaviridae; Coronavirus.

OX NCBI_TaxID=11133;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=90171910; PubMed=2155283;

RA Parker M.D., Yoo D., Cox G.J., Bahiuk L.A.;

RT "Primary structure of the S peplomer gene of bovine coronavirus and

surface expression in insect cells."

RL J. Gen. Virol. 71:263-270(1990).

CC -1- FUNCTION: THE PEPLIMER PROTEIN MEDIATES THE BINDING OF VIRIONS

CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: D00662; BAA00557.1; -.

DR PIR: A34147; VG1H0U.

DR InterPro: IPR002552; Corona_S2.

DR Pfam: PF01601; Corona_S2; 2.

KM Glycoprotein; Envelope protein; Transmembrane; Signal.

FT SIGNAL 1 17

FT CHAIN 18 1363

FT CHAIN 18 768

FT CHAIN 769 1363

FT DOMAIN 18 1311

FT TRANSMEM 1312 1328

FT DOMAIN 1329 1363

FT DOMAIN 1329 1346

FT CARBOHYD 59 59

FT CARBOHYD 133 133

FT CARBOHYD 198 198

FT CARBOHYD 359 359

FT CARBOHYD 437 437

FT CARBOHYD 649 649

FT CARBOHYD 676 676

FT CARBOHYD 696 696

FT CARBOHYD 714 714

FT CARBOHYD 739 739

FT CARBOHYD 788 788

FT CARBOHYD 937 937

FT CARBOHYD 1194 1194

FT CARBOHYD 1224 1224

FT CARBOHYD 1234 1234

FT CARBOHYD 1253 1253

FT CARBOHYD 1267 1267

FT CARBOHYD 1288 1288

SO SEQUENCE 1363 AA; 150869 MW; 6A6587B07A102B71 CRC64;

Best Local Similarity 53.8%; Pred. No. 81;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 1 GNAEHVASAVENA 13
Db 1138 GNGNHITSLVQNA 1150

Search completed: June 10, 2002, 15:29:42
Job time: 477 sec

Query Match

45.6%; Score 41; DB 1; Length 1363;

DR WPI; 1993-386472/48.
 XX
 PT Universal corona-virus vaccine - comprising a polypeptide contg.
 PT a conserved domain of coronavirus S protein or nucleic acid
 PT encoding it
 XX
 PS Claim 1; Page 26-27; 36pp; English.
 XX
 CC The polypeptide comprises a sequence which is highly conserved among
 CC coronaviruses and which is capable of eliciting an immune response.
 CC The polypeptide and nucleic acid can be used to produce vaccines
 CC which can be used to protect animals against different members of
 CC the coronavirus family, e.g. feline infectious peritonitis virus,
 CC feline enteric coronavirus; canine coronavirus; swine transmissible
 CC gastroenteritis coronavirus; bovine coronavirus; human coronavirus
 CC or avian infectious bronchitis virus.
 XX
 SQ Sequence 180 AA;

Query Match 45.6%; Score 41; DB 14; Length 180;
 Best Local Similarity 53.8%; Pred. No. 64;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GNAEHVASAVENA 13
 || | : | 1:11
 Db 112 gngnhlslvqna 124

Search completed: June 10, 2002, 15:21:07
 Job time: 167 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 15:22:23 ; Search time 35.61 seconds

(without alignments)
48.571 Million cell updates/sec

Title: US-09-647-522-2

Perfect score: 90

Sequence: 1 GNAEHVASAVENANRVNK 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	90	100.0	450	2 JC7371	toxin - jellyfish
2	45	50.0	547	2 AF3302	hypothetical prote
3	44	48.9	471	2 S30585	Vital replicase I
4	44	48.9	653	2 U01241	pac cutting enzyme
5	43	47.8	142	2 C41045	auxin-induced prot
6	43	47.8	194	2 T10939	thiamin-phosphate
7	43	47.8	221	2 AB3018	thiamin-phosphate
8	43	47.8	238	2 F98266	triose-phosphate
9	43	47.8	249	1 ISASTN	tylaxose-phosphate
10	43	47.8	297	1 IS8517	ABC transporter xy
11	43	47.8	397	2 S21780	ABC transporter xy
12	43	47.8	461	2 B90555	ABC transporter xy
13	43	47.8	676	2 C97775	acylamino-acid-rel
14	43	47.8	814	1 I39627	nicotine dehydroge
15	42	46.7	163	2 F64018	conserved hypotet
16	42	46.7	312	2 F86044	secreted protein E
17	42	46.7	312	2 B91198	EsPB protein (limp
18	42	46.7	478	2 AB3641	funarate hydratase
19	42	46.7	481	2 A90869	hypothetical prote
20	42	46.7	481	2 H85749	hypothetical prote
21	42	46.7	481	2 D64883	Aminobenzoyl-gluta
22	41.5	46.1	353	2 A82405	hypothetical prote
23	41.5	46.1	915	2 T09575	smoothenin - human
24	41	45.6	590	2 B81104	nitrate/nitrite se
25	41	45.6	590	2 C81911	nitrate/nitrite se
26	41	45.6	1235	1 VGIHMF	E2 glycoprotein pr
27	41	45.6	1324	1 VGIH59	E2 glycoprotein pr
28	41	45.6	1353	1 U02168	E2 glycoprotein pr
29	41	45.6	1361	2 S29998	surface protein -

30	41	45.6	1362	2 A37474	surface glycoprote
31	41	45.6	1363	1 VGIHNM	E2 glycoprotein pr
32	41	45.6	1363	1 VGIHQU	E2 glycoprotein pr
33	41	45.6	1363	1 VGIHFI	E2 glycoprotein pr
34	41	45.6	1363	1 VGIHL9	E2 glycoprotein pr
35	41	45.6	1363	1 VGIHVA	E2 glycoprotein pr
36	41	45.6	1363	1 VGIHLV	E2 glycoprotein pr
37	41	45.6	1363	1 S44240	surface protein -
38	41	45.6	1363	2 S44241	surface protein -
39	41	45.6	1376	1 VGIHJ2	E2 glycoprotein pr
40	41	45.6	1376	1 J01534	E2 glycoprotein pr
41	41	45.6	3890	2 C89921	hypothetical prote
42	40	44.4	75	2 E97819	hypothetical prote
43	40	44.4	231	2 H69407	conserved hypotet
44	40	44.4	311	2 E98240	ABC transporter ho
45	40	44.4	311	2 AG3045	dlxoygenase [limpor

ALIGNMENTS

RESULT 1
JC7371
toxin - jellyfish (Carybdea rastoni)
C:Species: Carybdea rastoni
C:Date: 17-Nov-2000 #sequence-revision 17-Nov-2000 #text-change 17-Nov-2000
C:Accession: JC7371; PC7094
R:Nagai, H.; Takawa, K.; Nakao, M.; Ito, E.; Miyake, M.; Noda, M.; Nakajima, T.
Biochem. Biophys. Res. Commun. 275, 582-588, 2000
A:Title: Novel proteinaceous toxins from the box jellyfish (sea wasp) Carybdea raston
A:Reference number: JC7371
A:Accession: JC7371
A:Molecule type: mRNA
A:Residues: 1-450 <NNG>
A:Cross-references: DDBJ:AB015878
A:Accession: PC7094
A:Molecule type: protein
A:Residues: 39-55;56-70;196-210;250-267;268-279;309-325;363-377;378-382 <NAG>
C:Comment: This protein, a member of bioactive protein, has hemolytic activity.
C:Keywords: hemolysis, inflammation, toxin

Query Match 100.0%; Score 90; DB 2; Length 450;
Best Local Similarity 100.0%; Pred. No. 7.1e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GNAEHVASAVENANRVNK 18
DB 250 GNAEHVASAVENANRVNK 267

RESULT 2
AF3302
hypothetical protein BME10404 [Imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence-revision 01-Feb-2002 #text-change 01-Feb-2002
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanov
; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3302
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-547 <KUR>
A:Cross-references: GB:AE008917; PIDN:ALU5185.1; PID:q17982308; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME10404
A:Map position: 1

A:Gene: th1E
A:Map position: linear chromosome

Query Match 47.8%; Score 43; DB 2; Length 221;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GNAEHVASAVEN 12
|||||
DB 94 GNAEALAEAVEN 105

RESULT 8
F98266
thiamin-phosphate pyrophosphorylase (AF180145) [Imported] - Agrobacterium tumefaciens (S
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
R:Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: F98266
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-238 <KUR>
A:Cross-references: GB:A007870; PIDN:AAK89656.1; PID:915159556; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_2166
A:Map position: linear chromosome

Query Match 47.8%; Score 43; DB 2; Length 238;
Best Local Similarity 75.0%; Pred. No. 17;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GNAEHVASAVEN 12
|||||
DB 111 GNAEALAEAVEN 122

RESULT 9
ISASTN
triose-phosphate isomerase (EC 5.3.1.1) - Emericella nidulans
N:Alternate names: triosephosphate mutase
C:Species: Emericella nidulans, Aspergillus nidulans
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jun-2000
R:McKnight, G.L.; O'Hara, P.J.; Parker, M.L.
Cell 46, 143-147, 1986
A:Title: Nucleotide sequence of the triosephosphate isomerase gene from Aspergillus nid
A:Reference number: A25502; MUID:86245050
A:Accession: A25502
A:Molecule type: DNA
A:Residues: 1-249 <MCK>
A:Cross-references: GB:D10019; GB:M13362; NID:q217920; PIRN:BA00908.1; PID:q217921
C:Comment: This enzyme catalyzes the interconversion of glyceraldehyde 1-phosphate and c
C:Genetics:
A:Introns: 13/2; 106/3; 131/3; 169/2; 241/1
C:Superfamily: triose-phosphate isomerase
C:Keywords: glucogenesis; glycolysis; homodimer; intramolecular oxidoreductase; isome
F:2-249/Product: triose-phosphate isomerase #status predicted <MKT>
F:94/166/Active site: His, Glu #status predicted

Query Match 47.8%; Score 43; DB 1; Length 249;
Best Local Similarity 47.1%; Pred. No. 18;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 GNAEHVASAVENANRN 17
|||||

DB 15 GNAESTTSIIKNLSAN 31

RESULT 10

I38517

synthaxin - human

C:Species: Homo sapiens (man)

C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 24-Sep-1999

C:Accession: I38517

R:Li, H.; Hodge, D.R.; Pel, G.K.; Seth, A.

Gene 143, 303-304, 1994

A:Title: Isolation and sequence analysis of the human synthaxin-encoding gene.

A:Reference number: I38517; MUID:94266173

A:Accession: I38517

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-297 <RES>

A:Cross-references: EMBL:007158; NID:q463906; PIDN:AAA20967.1; PID:q463907

C:Superfamily: synthaxin

Query Match 47.8%; Score 43; DB 2; Length 297;
Best Local Similarity 53.3%; Pred. No. 22;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 4 EHVASAVENANRNK 18
|||||

DB 257 EHVKTALENOKVKRK 271

RESULT 11

S21780

pacB protein - phage P1

C:Species: phage P1

C:Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 08-Oct-1999

C:Accession: S21780

R:Skorupski, K.; Pierce, J.C.; Sauer, B.; Sternberg, N.

J. Mol. Biol. 223, 977-989, 1992

A:Title: Bacteriophage P1 genes involved in the recognition and cleavage of the phage

A:Reference number: S21779; MUID:92167273

A:Accession: S21780

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-397 <SKO>

A:Cross-references: EMBL:M74046; NID:q215634; PIDN:AAA21723.1; PID:q215635

C:Genetics:

A:Start codon: GTG

C:Keywords: DNA binding

Query Match 47.8%; Score 43; DB 2; Length 397;
Best Local Similarity 69.2%; Pred. No. 30;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 6 VASAVENANRNK 18
|||||

DB 93 VSAAVENAKRNK 105

RESULT 12

B90555

ABC transporter xylose-binding lipoprotein [Imported] - Mycoplasma pulmonis (strain U

C:Species: Mycoplasma pulmonis

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001

C:Accession: B90555

R:Chamblaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer,

Nucleic Acids Res. 29, 2145-2153, 2001

A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p

A:Reference number: A99512; MUID:21267165; PMID:11353084

A:Accession: B90555

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-461 <KUR>

Page 4

Query Match 47.8%; Score 43; DB 22; Length 605;
Best Local Similarity 47.1%; Pred. No. 1.2e+02;
Matches 8; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 GNAEHVASAVENANRVN 17
DB 366 glfgnlangvnanrln 382

RESULT 11

AAB35466
ID AAB35466 standard; Protein; 312 AA.

XX AAB35466;

AC AAB35466;

XX 06-JUN-2001 (first entry)

DE Escherichia coli EspB #1.

XX Escherichia coli.

OS Escherichia coli.

XX Key Location/Qualifiers

FT Misc-difference 199 /note="encoded by AAT"

FT Misc-difference 203 /note="encoded by CCC"

FT Misc-difference 219 /note="encoded by AAA"

XX US6204004-B1.

XX 20-MAR-2001.

XX 21-MAR-1997; 97US-0821872.

XX 21-MAR-1997; 97US-0821872.

XX (UYMA-) UNIV MARYLAND BALTIMORE.

XX Kaper JB, Jarvis K;

XX WPI; 2001-256675/26.

XX N-PSDB; AAF58936.

XX Diagnosis of active infection by enterohemorrhagic Escherichia coli

XX comprises detecting antibodies to E. coli secreted protein EspA or EspB

XX Disclosure; Column 17-20; 19pp; English.

XX The present invention describes a method of diagnosing infection by

XX enterohemorrhagic Escherichia coli 0157:H7 by contacting a body fluid

XX sample with EspA or EspB and detecting the presence of antibodies. E.

XX coli infection causes haemorrhagic colitis and can lead to death or long

XX term liver damage. The present sequence is one version of the EspB

XX protein.

XX Note: The present sequence is stated as being the same as that given in

XX columns 7-10 in the specification (see AAB35471). However, the sequences

XX differ at three positions.

XX SQ Sequence 312 AA;

RESULT 12
ID ABB64054 standard; Protein; 378 AA.

XX ABB64054;

AC ABB64054;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 18954.

XX Drosophila: developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL08157.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions

XX Disclosure; SEQ ID NO 18954; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins

XX (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 378 AA;

XX Query Match 46.7%; Score 42; DB 22; Length 378;

XX Best Local Similarity 46.7%; Pred. No. 1e+02;

XX Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

XX QY 2 NAEHVASAVENANRV 16

XX DB 240 nsghatalekav 254

RESULT 13
ID AAU34565 standard; Protein; 481 AA.

XX AAU34565;

AC AAU34565;

XX 14-FEB-2002 (first entry)

DE E. coli cellular proliferation protein #146.

XX Antisense; prokaryotic cellular proliferation protein;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

antibiotic; antibacterial; drug design.

XX
OS Escherichia coli.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
PF 21-MAR-2000; 2000US-191078P.
XX
PR 23-MAY-2000; 2000US-206848P.
XX
PR 26-MAY-2000; 2000US-207727P.
XX
PR 23-OCT-2000; 2000US-242578P.
XX
PR 27-NOV-2000; 2000US-253625P.
XX
PR 22-DEC-2000; 2000US-257931P.
XX
PR 16-FEB-2001; 2001US-265308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI: 2001-611495/70.
DR N-PSDB; AAS52424.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 10158; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC Escherichia coli, and the encoded proteins. The prokaryotes used are
CC pneumonae, Pseudomonas deruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 481 AA;
SQ

Query Match 46.7%; Score 42; DB 22; Length 481;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 NAENVASAVENA 13
DB 39 saehlaaalea 50

RESULT 14
ID AAR43880 standard; protein; 179 AA.
XX
XX AAR43880;
XX
XX 25-MAY-1994 (first entry)
XX
XX C-terminal portion of BCV spike protein (Strain Bcve2).
XX

Coronavirus; spike protein; S protein; vaccine; BCV;
XX bovine coronavirus.
XX
XX Bovine coronavirus.
XX
XX WO9323421-A.
XX
XX 25-NOV-1993.
XX
XX 07-MAY-1993; 93WO-US04365.
XX
XX 08-MAY-1992; 92US-0882171.
XX
XX (SMK) SMITHKLINE BEECHAM CORP.
XX
XX Jones EV, Klepfer S, Miller TJ, Reed AP;
XX
XX WPI: 1993-386472/48.
DR
XX Universal corona-virus vaccine - comprising a polypeptide contg.
PT a conserved domain of coronavirus S protein or nucleic acid
PI encoding it
XX
XX Claim 1; Page 24-25; 36pp; English.
XX
XX The polypeptide comprises a sequence which is highly conserved among
CC coronaviruses and which is capable of eliciting an immune response.
CC The polypeptide and nucleic acid can be used to produce vaccines
CC which can be used to protect animals against different members of
CC the coronavirus family, e.g. feline infectious peritonitis virus,
CC feline enteric coronavirus; canine coronavirus; swine transmissible
CC gastroenteritis coronavirus; bovine coronavirus; human coronavirus
CC or avian infectious bronchitis virus.
XX
XX Sequence 179 AA;
SQ

Query Match 45.6%; Score 41; DB 14; Length 179;
Best Local Similarity 53.8%; Pred. No. 64;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 GNAENVASAVENA 13
DB 111 gngnhlslvqna 123

RESULT 15
ID AAR43883 standard; protein; 180 AA.
XX
XX AAR43883;
XX
XX 25-MAY-1994 (first entry)
XX
XX C-terminal portion of MHV spike protein (Strain Mhve2a59).
DE
XX Coronavirus; spike protein; S protein; vaccine; FIV;
XX mouse hepatitis virus.
XX
XX Mouse hepatitis virus.
XX
XX WO9323421-A.
XX
XX 25-NOV-1993.
XX
XX 07-MAY-1993; 93WO-US04365.
XX
XX 08-MAY-1992; 92US-0882171.
XX
XX (SMK) SMITHKLINE BEECHAM CORP.
XX
XX Jones EV, Klepfer S, Miller TJ, Reed AP;
XX

New recombinant plasmids to integrate in *Aspergillus* chromosomal DNA - useful for expressing higher eucaryotic genes esp. when plasmid contains transcriptional promoter, eg DNA encoding alcohol dehydrogenase

Disclosure; Fig 1; 79pp; English.

This sequence is used for expressing higher eucaryotic genes in *Aspergillus* through the use of a recombinant plasmid capable of integration into the chromosome DNA of *Aspergillus*. See also AANT0814-16 and AAB70497-98.

Sequence 250 AA:

Query Match 47.8%; Score 43; DB 8; Length 250;
Best Local Similarity 47.1%; Pred. No. 44;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 GNAEHVSAVENANRVN 17
| | | | | : : : | |
Db 15 gnaesttslknlnsan 31

RESULT 9
AAY07078
ID AAY07078 standard; Protein: 297 AA.
AC AAY07078;
XX
DT 02-JUL-1999 (first entry)
XX
DE Renal cancer associated antigen precursor sequence.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human; breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer; prostate cancer.
XX
OS Homo sapiens.
XX
PN M09904265-A2.
XX
PD 28-JAN-1999.
XX
PE 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfeundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Disclosure; Page 484; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the

Interaction between the agent and the NAM or the expression product as a
determination of the disorder. The products and methods can be used in
the diagnosis, monitoring, research, or treatment of conditions
characterised by the expression of various cancer associated antigens.
The invention provides nucleic acid sequences and encoded polypeptides
which are cancer associated antigen precursors expressed in human breast
cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
lung cancer.

Sequence 297 AA:

Query Match 47.8%; Score 43; DB 20; Length 297;
Best Local Similarity 53.3%; Pred. No. 54;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 4 EHVASAVENANRVN 18
| | | : | : | : |
Db 257 ehvktalengkkvrk 271

RESULT 10
ABB67887
ID ABB67887 standard; Protein: 605 AA.
AC ABB67887;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 30453.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX
OS Drosophila melanogaster.
XX
PN W0200171042-A2.
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PDB; ABL11990.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 30453; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB101845), expressed DNA
CC sequences (AB101840-AB101845) and the encoded proteins
CC (AB101840-AB101845).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pt_sequences.
XX
SQ Sequence 605 AA;

XX 14-APR-1999; 99US-0291902.
PR 13-APR-2000; 2000US-0548449.
XX
PA (MUSC-) MUSC FOUND RES DEV.
XX (UNYF-) UNIV PENN STATE.
XX
PI Norris J, Clawson G, Westwater C, Schofield D, Schmidt M, Hoel B;
PI Dolan J, Pan W;
XX
XX WPI: 2000-638570/61.
DR N-PSDB: AAA95507.
XX
XX
PI New recombinant nucleic acid comprising a nucleotide sequence encoding
PI one or more toxic agents operably linked to a pathogen specific or
PI tissue specific promoter is useful for inhibiting replication of a
PI pathogen -
XX
XX
PS Disclosure: Fig 12; 113pp; English.
XX
XX
CC The present sequence comprises the minimal PI pac site. It was used in
CC the construction of the pathogen and tissue-specific toxic agents of the
CC invention. These agents may be antisense sequences, ribozymes or toxic
CC proteins. These can be used in the treatment of cancer and infections
CC including hepatitis, herpes, malaria, bacterial meningitis, typhoid fever
CC and tuberculosis.
XX
XX
SQ Sequence 59 AA:

Query Match 47.8%; Score 43; DB 21; Length 59;
Best Local Similarity 69.2%; Pred. No. 8.7;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 6 VASAVENANRVNK 18
|::| | | | | | | | | |
DB 33 vsaavenakrqnk 45

RESULT 7

AAM52135
ID AAM52135 standard; protein; 59 AA.

XX
AC AAM52135;

DT 05-FEB-2002 (first entry)

XX
DE Minimal PI pac site sequence protein.

XX
KM Infection: antisense RNA; ribozyme; DNAzyme; antiviral; gene therapy;
KM papilloma virus; hepatitis B virus; cytotoxic; cytostatic; wart;
KM cervical dysplasia; cervical carcinoma; carcinoma; laryngeal papilloma;
KM bacteriophage PI; pac site.

XX
OS Bacteriophage PI.

XX
FH Key Location/Qualifiers

FT Misc-difference 27
/note= "Encoded by CAA"

XX
PN WO200179524-A2.

XX
PD 25-OCT-2001;

PF 13-APR-2001; 2001WO-US12130.

PR 13-APR-2000; 2000US-0548449.

PR 07-DEC-2000; 2000US-251810P.

XX
PA (UNYF-) UNIV SOUTH CAROLINA.
PA (PENN-) PENN STATE RES FOUND.

XX
PI Norris JS, Clawson GA, Westwater C, Schofield D, Schmidt MG;

PI Hoel B, Dolan J, Pan W;
XX
XX WPI: 2001-607700/69.
DR N-PSDB: ABA02577.
XX

PT Novel nucleic acid for the treatment of papilloma or hepatitis virus
PT induced conditions comprises a catalytic region which produces a
PT cytotoxic or cytosstatic effect in the infected cell -
XX
XX
PS Examples: Fig 12; 143pp; English.

XX
XX
CC The invention relates to the discovery, identification and
CC characterisation of toxic agents lethal to pathogens and methods for
CC targeting such toxic agents to a pathogen or pathogen infected cells in
CC order to treat and/or eradicate the infection, in particular the
CC invention relates to at least one nucleic acid molecule, which
CC specifically hybridises to mRNA encoding at least one viral protein
CC associated with the transformation or plasmid copy number control, which
CC hybridises to a viral polyadenylation signal or a core, pre core or
CC polymerase encoding sequence. Specifically, the invention relates to the
CC delivery of one or more toxic gene products, antisense RNAs, ribozymes,
CC DNAzymes or a combination thereof. The nucleic acids have antiviral
CC activity and can be used in gene therapy. They are useful for the
CC treatment of papilloma or hepatitis virus induced conditions and can
CC produce a cytotoxic or cytosstatic effect in papillomavirus or hepatitis B
CC infected cells. The papilloma virus induced condition is selected from
CC warts, cervical dysplasia, cervical carcinoma, carcinoma in situ and
CC laryngeal papilloma. The present sequence is that of a modified
CC bacteriophage PI pac site sequence rendered defective due to several
CC silent mutations.
XX
XX
SQ Sequence 59 AA:

Query Match 47.8%; Score 43; DB 22; Length 59;
Best Local Similarity 69.2%; Pred. No. 8.7;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 6 VASAVENANRVNK 18
|::| | | | | | | | | |
DB 33 vsaavenakrqnk 45

RESULT 8

AAP70496
ID AAP70496 standard; protein; 250 AA.

XX
AC AAP70496;

DT 27-FEB-1991 (first entry)

XX
DE Aspergillus nidulans tp1A genomic clone.

XX
KM tp1A gene; triose-phosphate-isomerase; enzyme; promoter.

XX
OS Aspergillus nidulans.

XX
PN WO8704464-A.

XX
PD 30-JUL-1987.

PF 15-JAN-1987; 87WO-US00119.

PR 13-JAN-1987; 87US-0946873.

PR 17-JAN-1986; 86US-0820519.

XX
PA (ZYMO-) ZYMOGENETICS INC.

XX
PI McKnight GL, Upshall A;

XX
DR WPI: 1987-221265/31.

XX
DR N-PSDB: AAN70813.

XX

```

RESULT 4
AAR13992
ID AAR13992 standard; Protein; 493 AA.
XX
AC AAR13992:
XX
DE 05-DEC-1991 (first entry)
XX
DE P. falciparum sporozoite antigen fusion polypeptide.
XX
KM Vaccine; NXY gene; malaria; antibodies; purification; affinity.
XX
OS Plasmodium falciparum.
XX
FH Key
FT Region
FT 1..21
FT /label=A
FT /note="affinity peptide including 6 His residues"
FT 22..483
FT /label=B
FT /note="amino acids 1-462 of the peptide in AAR13991"
FT 484..493
FT /label=C
FT /note="vector-encoded_peptide"
XX
PN EP447956-A.
XX
PD 25-SEP-1991.
XX
PF 14-MAR-1991; 91EP-0103920.
XX
PR 23-MAR-1990; 90CH-0000970.
XX
PA (HOFF ) HOFFMANN-LA ROCHE AG.
XX
PI Certa U, Guttinger M, Matile H;
XX
DR WPI; 1991-282989/39.
XX
PT New antimalarial polypeptide(s) - corresp. to specific epitope(s)
PT of the Plasmodium falciparum sporozoite antigen, and are useful
PT as vaccines or to prepare antigens
XX
PS Claim 2; Page 18; 36pp; German.
XX
CC Vector pDS56/RBSII. 6xHis is a derivative of pDS56/RBS (EP-282042).
CC It comprises an additional sequence encoding six histidine residues.
CC Into this vector may be ligated a 1400 bp AseI fragment of NXY
CC (see AAQ13727) to yield pDS-NXY. E. coli SG13009 (pUHA1) transformed
CC with pDS-NXY produces a 69 kD fusion protein comprising the NXY
CC fragment N-terminally linked to an affinity peptide contg. 6 His
CC units, and C-terminally to a vector-derived sequence.
CC Fragment B corresponds to amino acids 1-462 of the N-terminal of
CC the sporozoite antigen. The affinity peptide is used for the
CC purification of the antigen. The polypeptide is useful for the
CC prodn. of antimalarial vaccines.
CC See also AAQ13727-8.
XX
SQ Sequence 493 AA;

Query Match 48.9%; Score 44; DB 12; Length 493;
Best Local Similarity 52.9%; Pred. No. 66;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
OY 2 NAEHVASAVENANRVNK 18
DB 26 navhkhlnavdkvnavnk 42

```

```

XX
AC AAR13991;
XX
DE 05-DEC-1991 (first entry)
XX
DE Plasmodium falciparum sporozoite antigen N-terminal.
XX
KM Vaccine; NXY gene; malaria; antibodies.
XX
OS Plasmodium falciparum.
XX
PN EP447956-A.
XX
PD 25-SEP-1991.
XX
PF 14-MAR-1991; 91EP-0103920.
XX
PR 23-MAR-1990; 90CH-0000970.
XX
PA (HOFF ) HOFFMANN-LA ROCHE AG.
XX
PI Certa U, Guttinger M, Matile H;
XX
DR WPI; 1991-282989/39.
XX
DR N-PSDB; AAQ13727.
XX
PT New antimalarial polypeptide(s) - corresp. to specific epitope(s)
PT of the Plasmodium falciparum sporozoite antigen, and are useful
PT as vaccines or to prepare antigens
XX
PS Claim 1; Page 16; 36pp; German.
XX
CC Polypeptides corresp. in at least one specific epitope with the
CC P. falciparum sporozoite are useful in the prodn. of antimalarial
CC vaccines. The N-terminal has a mol.wt. of 81.281, the total protein
CC has a mol.wt. above 200 kD. Potential glycosylation sites are at
CC position 32, 260, 308, 323, 329, 344, 362, 365, 377, 380, 387, 388,
CC 394, 398, 399, 406, 414, 537, 554, 659, 684, 693, 702, and 705.
CC See also AAQ13728.
XX
SQ Sequence 713 AA;

Query Match 48.9%; Score 44; DB 12; Length 713;
Best Local Similarity 52.9%; Pred. No. 1e+02;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
OY 2 NAEHVASAVENANRVNK 18
DB 5 navhkhlnavdkvnavnk 21

RESULT 6
AAB27206
ID AAB27206 standard; Protein; 59 AA.
XX
AC AAB27206;
XX
DE 27-FEB-2001 (first entry)
XX
DE Bacteriophage delivery plasmid minimal p1 pac site peptide.
XX
DE Bacteria; infection; drug-resistant pathogen; cancer; typhoid fever;
XX bacterial meningitis; tuberculosis; antisense strand; ridozyme;
XX toxic protein.
XX
OS Unidentified.
XX
PN WO200061804-A1.
XX
PD 19-OCT-2000.
XX
PF 14-APR-2000; 2000WO-US10229.

```

PS Claim 5; Page 22-23; 32pp; Japanese.

CC This invention describes a novel protein which has hemolytic activity,
 CC blood platelet agglutination activity and a molecular weight of about
 CC 50,000 Da (by SDS-PAGE). The protein, homologs and analogs of the
 CC invention can be used as a novel approach to developing drugs useful for
 CC the treatment of jelly fish stings, pharmaceuticals with blood platelet
 CC agglutination activity, pesticides by use of the hemolytic activity, and
 CC in the study of the hemolytic mechanism. AA13648-13650 represent
 CC fragments of the hemolytic protein described in the invention
 CC
 XX
 SQ Sequence 18 AA;

Query Match 100.0%; Score 90; DB 20; Length 18;
 Best Local Similarity 100.0%; Pred. No. 6.9e-08;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GNAEHVASAVENANRVNK 18
 |||||
 DB 1 gnaehvasavenanrvnk 18

RESULT 2

AA13651
 ID AA13651 standard; Protein; 450 AA.

XX AA13651;

DT 06-JAN-2000 (first entry)

DE C. rastonii hemolytic protein.

KW Hemolytic protein; blood platelet agglutination; drug development;
 KW treatment; sting; jellyfish; pharmaceutical; pesticide.

OS Carybdea rastonii.

PN W09950284-A1.

PD 07-OCT-1999.

PF 30-MAR-1999; 99WO-JP01607.

PR 01-APR-1998; 98JP-0088569.

XX (SUNR) SUNTORY LTD.

PI Nagai H, Nakajima T;

DR WPI, 1999-580740/49.

DR N-PSDB; AA23610.

PT Protein with hemolytic activity, useful for drugs treating jelly fish
 PT stings, pharmaceuticals with blood platelet agglutination activity,
 PT pesticides by use of the hemolytic activity, and study of the hemolytic
 PT mechanism

PS Claim 4; Page 26-27; 32pp; Japanese.

XX This invention describes a novel protein which has hemolytic activity,
 CC blood platelet agglutination activity and a molecular weight of about
 CC 50,000 Da (by SDS-PAGE). The protein, homologs and analogs of the
 CC invention can be used as a novel approach to developing drugs useful for
 CC the treatment of jelly fish stings, pharmaceuticals with blood platelet
 CC agglutination activity, pesticides by use of the hemolytic activity, and
 CC in the study of the hemolytic mechanism. This sequence represents the
 CC hemolytic protein described in the invention
 CC
 XX
 SQ Sequence 450 AA;

Query Match 100.0%; Score 90; DB 20; Length 450;

Best Local Similarity 100.0%; Pred. No. 2.6e-06;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GNAEHVASAVENANRVNK 18
 |||||
 DB 250 gnaehvasavenanrvnk 267

RESULT 3

AAW75903
 ID AAW75903 standard; peptide; 19 AA.

XX AAW75903;

DT 11-NOV-1998 (first entry)

DE Novel foam protein 1 for beer head retention.

KW Foam protein; beer head retention; monoclonal antibody; immunoassay;
 KW ELISA; brewing; beer stability; beer making; stabiliser; silica gel;
 KW tannic acid; fermentation.

OS Synthetic.

OS Hordeum sp.

PN EP663153-A2.

PD 09-SEP-1998.

PF 05-MAR-1998; 98EP-0103922.

PR 04-DEC-1997; 97JP-0334229.

PR 07-MAR-1997; 97JP-0053249.

XX (SUNR) SUNTORY LTD.

PI Ishibashi Y, Kakui T, Nakatani K, Terano Y;

DR WPI, 1998-458799/40.

PT New foam proteins crucial for head retention of beer - recognised
 PT by monoclonal antibodies used in immunoassays to assay foam protein
 PT content, head retention and raw materials of beer and to select
 PT stabiliser for beer

PS Claim 2; Page 5; 17pp; English.

XX Sequences shown in AAW75903 to AAW75908 are novel foam proteins with a
 CC molecular weight between 40-48 kd, as determined by Western blot
 CC analysis. These foam proteins are crucial for beer head retention.
 CC Monoclonal antibodies that specifically recognise any of these foam
 CC proteins are used in immunoassays (preferably ELISAs) for determining
 CC the foam protein content of beer during brewing, for determining the
 CC head retention or head retention stability of beer, for evaluating raw
 CC materials for beer making, and for selecting stabilisers (preferably
 CC silica gel or tannic acid) for beer. The immunoassays allow
 CC investigation of raw materials and stability during fermentation rather
 CC than after and they are an improvement on prior art assays involving
 CC polyclonal antibodies and allow detection of minute amounts of
 CC ingredient-specific foam proteins.

XX Sequence 19 AA;

Query Match 54.4%; Score 49; DB 19; Length 19;
 Best Local Similarity 100.0%; Pred. No. 0.27;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 AVENANRVNK 18
 |||||
 DB 1 avenanrvnk 10

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Choisme N., Robert C., Brotlier P., Wincker P., Catolico L.,
 RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Rudd S.,
 RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.,
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL049659; CAB41155.1; -
 DR InterPro: IPR003347; JmjC.
 DR InterPro: IPR003349; JmjN.
 DR Pfam: PF02373; JmjC; 1.
 DR Pfam: PF02375; JmjN; 1.
 DR Pfam: PF00096; zf-C2H2; 4.
 DR SMART: SM00545; JmjN; 1.
 DR SMART: SM00355; ZNF_C2H2; 4.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 3.
 KW DNA-binding; Metal-binding; Zinc-finger.
 SQ SEQUENCE 1360 AA; 152628 MW; A45A5801285EBA16 CRC64;

Query Match 52.8%; Score 38; DB 10; Length 1360;
 Best Local Similarity 41.7%; Pred. No. 4.5e+02;
 Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 OY 1 GEIOTKPPDRVGO.12
 Db 1195 GSIETKPKKIGK 1206

Search completed: June 10, 2002, 15:29:16
 Job time: 486 sec

KW	Hypothetical protein; complete proteome.
SQ	SEQUENCE 476 AA; 50680 MW; 935A0C47A73EE6D4 CRC64;

05 *Arabidopsis thaliana* (Mouse-ear cress).
0C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta

RA Kinasli H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AL117385; CAB55654.1; -
 SQ SEQUENCE 425 AA; 44529 MW; EA0E711A361E92D CRC64;

Query Match 54.2%; Score 39; DB 2; Length 425;
 Best Local Similarity 77.8%; Pred. No. 88;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 OY 4 QTKPDYVGO 12
 Db 245 QQKPDRLGQ 253

RESULT 9
 O9W784 PRELIMINARY; PRT; 218 AA.
 AC O9W784;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE THYROID HORMONE RECEPTOR BETA (FRAGMENT).
 GN TR-BETA.
 OS Salmo salar (Atlantic salmon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OX NCBI_TaxID=8030;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rogers S.A., Sweeney G.E., Wigham T.;
 RT "Isolation of tr-alpha and tr-beta genes from Atlantic salmon.";
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 DR EMBL; AF146777; AAD38691.1; -
 DR HSSP; P10828; ZML.
 DR InterPro; IPR001628; zf-C4.
 DR Pfam; PF00105; zf-C4; 1.
 DR PRINTS; PR000447; STROIDFINGER.
 DR SMART; SM00399; ZNF_C4; 1.
 DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
 KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
 FT NON_TER 218
 RT SEQUENCE 218 AA; 25155 MW; 24802667DC3E843B CRC64;
 SQ SEQUENCE 218 AA; 25155 MW; 24802667DC3E843B CRC64;
 Query Match 52.8%; Score 38; DB 13; Length 218;
 Best Local Similarity 53.8%; Pred. No. 66;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 OY 1 GEIQTKPDYVGO 13
 Db 176 GVKETKPEDIGQA 188

OX NCBI_TaxID=8022;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21094002; PubMed=11174854;
 RA Marchand O., Safi R., Escriva H., Van Rompaey E., Prunet P.,
 RA Laudet V.;
 RT "Molecular cloning and characterization of thyroid hormone receptors
 in teleost fish.";
 RL J. Mol. Endocrinol. 26:51-65(2001).
 DR EMBL; AF302246; AAL06726.1; -
 KW Receptor.
 FT NON_TER 1
 RT NON_TER 1
 FT NON_TER 229
 SQ SEQUENCE 229 AA; 26357 MW; 68A25F1428EAE964 CRC64;

Query Match 52.8%; Score 38; DB 13; Length 229;
 Best Local Similarity 53.8%; Pred. No. 69;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 GEIQTKPDYVGO 13
 Db 127 GVKETKPEDIGQA 139

RESULT 11
 O9PM69 PRELIMINARY; PRT; 368 AA.
 AC O9PM69;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PUTATIVE ATP/GTP-BINDING PROTEIN (MRP PROTEIN HOMOLOGY).
 GN MRP OR Cj1606.
 GN Campylobacter jejuni.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 CC Campylobacter.
 OX NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCCTC 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Kelley J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
 RA Jagsels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,
 RA Whitehead S., Barrrell B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 reveals hypervariable sequences.";
 RL Nature 403:665-668(2000).
 DR EMBL; AL139079; CAB73594.1; -
 DR InterPro; IPR002744; DUF59.
 DR InterPro; IPR000392; NitrogenaseII.
 DR Pfam; PF01883; DUF59; 1.
 DR Pfam; PF00142; fer4_NiH; 1.
 KW Complete proteome.
 FT NON_TER 368 AA;
 RT SEQUENCE 368 AA; 40119 MW; C6AB8AA98A6F7036 CRC64;
 SQ SEQUENCE 368 AA; 40119 MW; C6AB8AA98A6F7036 CRC64;
 Query Match 52.8%; Score 38; DB 16; Length 368;
 Best Local Similarity 58.3%; Pred. No. 11e+02;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 OY 1 GEIQTKPDYVGO 12
 Db 146 GERTQPEVVGQ 157

RESULT 12
 O90Y22 PRELIMINARY; PRT; 396 AA.
 AC O90Y22;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)

```

09GYL4
ID 09GYL4 PRELIMINARY: PRT: 997 AA.
AC 09GYL4:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHEtical 112.0 KDA PROTEIN.
GN R04E5.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA MEDLINE=99069613; PubMed=9851916;
RX None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Miller N.;
RT "The sequence of C. elegans cosmid R04E5.";
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: U41538; AAG00010.1;
DR InterPro: IPR002965; P_fich_extensions.
DR PRINTS: PR01217; PRICEXTENSN.
KV Hypothetical protein.
SQ SEQUENCE 997 AA; 111954 MW; F1620378F0D9DB6 CRC64;

Query Match 55.6%; Score 40; DB 5; Length 997;
Best Local Similarity 60.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 4 QTRPRVGOA 13
DB 985 EARPRIQGA 994

RESULT 7
O9VNS3 PRELIMINARY: PRT: 300 AA.
AC O9VNS3:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE CG12588 PROTEIN.
GN CG12588.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abtill J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

```

```

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck R.J., Brokstein P., Brothier P.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Cantor A., Chandra I.,
RA Cherry J.M., Caley S., Dahke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003605; AAF52097.1;
DR FLYbase: FBgn0037272; CG12588.
SQ SEQUENCE 300 AA; 34133 MW; 4B6B5FA0E90ED10 CRC64;

```

```

Query Match 54.2%; Score 39; DB 5; Length 300;
Best Local Similarity 61.5%; Pred. No. 61;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 EIORKPRVGOAT 14
DB 123 KMOTKVDVAGQIT 135

RESULT 8
O9RL20 PRELIMINARY: PRT: 425 AA.
AC O9RL20:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE PUTATIVE SECRETED PROTEIN.
GN SC5G9.06.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,

```


DT 01-AUG-1998 (TREMblrel. 07, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE NUCLEOPORIN NUP153 HOMOLOG (FRAGMENT).
 OS Xenopus laevis (African clawed frog).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 CC Xenopodidae; Xenopus.
 NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98198465; PubMed=95351546;
 RA Shah S., Tugendreich S., Forbes D.J.;
 RT "Major Binding Sites for the Nuclear Import Receptor Are the Internal
 RT Nucleoporin Nup153 and the Adjacent Nuclear Filament Protein Tpr.";
 RL J. Cell Biol. 141:31-49(1998).
 DR EMBL: AF045567; AAC41273.1; -
 DR InterPro: IPR001876; Znf-Ranbp.
 DR Pfam: PF00641; Zf-Ranbp; 5.
 DR SMART: SM00547; Znf-RBZ; 5.
 KW Porln.
 FT NON_TER
 SO SEQUENCE 1219 AA; 123248 MW; 24EB9F96683656F6 CRC64;

Query Match 61.1%; Score 44; DB 13; Length 1219;
 Best Local Similarity 57.1%; Pred. No. 34;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 GEIOTKPDVGVGA 14
 Db 639 GSAGTKPDTLSQST 652

RESULT 3
 Q9BOE4 PRELIMINARY; PRT; 197 AA.
 AC Q9BOE4;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE STROMAL PROTEIN ASSOCIATED WITH THYMII AND LYMPH NODES SHORT
 DE ISOFORM.
 GN 1700021K02RIK OR SPATIAL.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 SCID; TISSUE=THYMUS;
 RX MEDLINE=21040340; PubMed=11196687;
 RA Flomerfelt F.A., Kim M.G., Schwartz R.H.;
 RT "Spatial, a gene expressed in thymic stromal cells, depends on three-
 RT dimensional thymus organization for its expression.";
 RL Genes Immun. 1:391-401(2000).
 DR EMBL: AF257503; AAG37104.1; -
 DR MGD: MGI:1923820; 1700021K02RIK.
 SO SEQUENCE 197 AA; 22180 MW; 3287375B23E0656F CRC64;

Query Match 59.7%; Score 43; DB 11; Length 197;
 Best Local Similarity 75.0%; Pred. No. 7.5;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 EIQTKPDVGVGA 13
 Db 175 EIQKPDVGVGA 186

RESULT 4
 Q9BOE5 PRELIMINARY; PRT; 231 AA.
 AC Q9BOE5;

DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE STROMAL PROTEIN ASSOCIATED WITH THYMII AND LYMPH NODES LONG
 DE ISOFORM.
 GN 1700021K02RIK OR SPATIAL.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 SCID; TISSUE=THYMUS;
 RX MEDLINE=21040340; PubMed=11196687;
 RA Flomerfelt F.A., Kim M.G., Schwartz R.H.;
 RT "Spatial, a gene expressed in thymic stromal cells, depends on three-
 RT dimensional thymus organization for its expression.";
 RL Genes Immun. 1:391-401(2000).
 DR EMBL: AF257502; AAG37103.1; -
 DR MGD: MGI:1923820; 1700021K02RIK.
 SO SEQUENCE 231 AA; 25812 MW; B99DB8A19BAD71B CRC64;

Query Match 59.7%; Score 43; DB 11; Length 231;
 Best Local Similarity 75.0%; Pred. No. 8.9;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 EIQTKPDVGVGA 13
 Db 209 EIQKPDVGVGA 220

RESULT 5
 O18464 PRELIMINARY; PRT; 337 AA.
 ID O18464;
 AC O18464;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE HYPOHETICAL 35.6 KDA PROTEIN PRECURSOR.
 GN HMEGFL-1.
 OS Herdmania momus.
 CC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
 CC Stolidobranchia; Pyuridae; Herdmania.
 NCBI_TaxID=7733;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CURVATA;
 RA Arnold J.M., Eri R., Lavin M.F.;
 RT "A novel gene containing multiple EGF-like motifs transiently
 RT expressed in the papillae of the ascidian tadpole larvae.";
 RL Dev. Dyn. 0:0-0(1997).
 DR EMBL: U82540; AAB67704.1; -
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002919; TIL.
 DR Pfam: PF01826; TIL; 4.
 DR SMART: SM00181; EGF; 4.
 KW Signal; Hypothetical protein.
 FT CHAIN 1 18
 FT SIGNAL 1 18
 SO SEQUENCE 337 AA; 35617 MW; 55CA0924118DBFC6 CRC64;

Query Match 55.6%; Score 40; DB 5; Length 337;
 Best Local Similarity 61.5%; Pred. No. 45;
 Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GEIOTKPDVGVGA 13
 Db 270 GNFQCPDVRVGA 282

RESULT 6

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 15:21:10 ; Search time 58.44 Seconds
(without alignments)
41.443 Million cell updates/sec

Title: US-09-647-522-1

Perfect score: 72

Sequence: 1 GEIOTKPDVGOAT 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_19.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	100.0	450	5	09GV72
2	44	61.1	1219	13	073749
3	43	58.7	197	11	09E024
4	43	58.7	231	11	09E025
5	40	55.6	337	5	018464
6	40	55.6	997	5	09GYL4
7	39	54.2	300	5	09VNS3
8	39	54.2	425	2	09RL20
9	38	52.8	218	13	09W784
10	38	52.8	229	13	09V127
11	38	52.8	368	16	09PM69
12	38	52.8	396	13	09V122
13	38	52.8	476	16	055723
14	38	52.8	623	17	027730
15	38	52.8	1360	10	09STM3
16	37	51.4	214	17	029456

17	37	51.4	241	3	09P707	09P707 schizosacch
18	37	51.4	279	16	09B673	09B673 rhizobium 1
19	37	51.4	298	2	09FBJ3	09FBJ3 streptomyce
20	37	51.4	339	2	032644	032644 helicobacte
21	37	51.4	339	16	025067	025067 helicobacte
22	37	51.4	339	16	09ZME1	09ZME1 helicobacte
23	37	51.4	390	11	09B663	09B663 rhizobium 1
24	37	51.4	457	11	09Z243	09Z243 mus musculu
25	37	51.4	601	2	09X8H7	09X8H7 streptomyce
26	37	51.4	601	12	0914M3	0914M3 suifolobus
27	37	51.4	803	10	09LMB7	09LMB7 chenopodium
28	37	51.4	968	12	0911G8	0911G8 white spot
29	37	51.4	6146	2	093HJ5	093HJ5 streptomyce
30	36.5	50.7	363	10	09LIV3	09LIV3 oryza sativ
31	36	50.0	152	2	093A39	093A39 synechococc
32	36	50.0	171	2	054205	054205 streptomyce
33	36	50.0	172	2	09K3M0	09K3M0 streptomyce
34	36	50.0	192	3	09Y806	09Y806 schizosacch
35	36	50.0	206	2	09S2W5	09S2W5 streptomyce
36	36	50.0	255	5	09TX14	09TX14 caenorhabdi
37	36	50.0	384	2	005492	005492 bacillus ce
38	36	50.0	389	2	087861	087861 streptomyce
39	36	50.0	406	2	09R657	09R657 bacillus ce
40	36	50.0	479	5	044505	044505 caenorhabdi
41	36	50.0	569	17	029266	029266 archaeglob
42	36	50.0	625	3	09UT17	09UT17 schizosacch
43	36	50.0	628	16	09UX41	09UX41 neisseria m
44	36	50.0	631	16	09K1G0	09K1G0 neisseria m
45	36	50.0	634	16	09KPE2	09KPE2 vibrio chol

ALIGNMENTS

RESULT 1
ID 09GV72 PRELIMINARY; PRT; 450 AA.
AC 09GV72:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE TOXIN-1.
GN CRT-1.
OS Carybdea rastoni.
OC Eukaryota; Metazoa; Cnidaria; Cubozoa; Cubomedusae; Carybdeidae.
CC Carybdea.
CX NCBI_TaxID=78582;
RN [1].
RP SEQUENCE FROM N.A.
RX MEDLINE=20422301; PubMed=10964707;
RA Nagai H., Takuwa K., Nakao M., Ito E., Miyake M., Noda M.,
RA Nakajima T.,
RT "Novel proteinaceous toxins from the box jellyfish (sea wasp) Carybdea
rastoni".
RL Blochem. Biophys. Res. Commun. 275:582-588(2000).
DR EMBL: AB015878; BAB12728.1; .
SQ SEQUENCE 450 AA; 49392 MW; CD393CF25BED2FD CRC64;

Query Match 100.0%; Score 72; DB 5; Length 450;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEIOTKPDVGOAT 14
DB 56 GEIOTKPDVGOAT 69
RESULT 2
ID 073749 PRELIMINARY; PRT; 1219 AA.
AC 073749:
DT 01-AUG-1998 (TREMBLrel. 07, Created)

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X61236; CAA43554.1; -
DR PIR: S19052; S19052.
DR SGD: S0002557; NDM1.
DR InterPro: IPR001849; PH.
DR Pfam: PF00169; PH; 1.
DR SMART: SM00233; PH; 1.
DR PROSITE: PSS0003; PH_DOMAIN; 1.
KW Repeat.
FT DOMAIN 593 1384 12.5 X TANDEM REPEATS.
FT REPEAT 593 656 1.
FT REPEAT 727 727 2.
FT REPEAT 728 798 3.
FT REPEAT 799 862 4.
FT REPEAT 863 926 5.
FT REPEAT 927 990 6.
FT REPEAT 991 1054 7.
FT REPEAT 1055 1118 8.
FT REPEAT 1119 1182 9.
FT REPEAT 1183 1246 10.
FT REPEAT 1247 1310 11.
FT REPEAT 1311 1374 12.
FT REPEAT 1375 1384 13. (INCOMPLETE).
FT DOMAIN 2573 2683 PH.
SQ SEQUENCE 2748 AA; 313202 MW; B2FBD67C9F6211AE CRC64;

Query Match 50.0%; Score 36; DB 1; Length 2748;
Best Local Similarity 58.3%; Pred. No. 3.2e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 IQTKPRVQAT 14
: 111:1111
DB 2441 LTRKDKRGAT 2452

RESULT 14
THB_PAROL STANDARD; PRT; 395 AA.
AC Q91279;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Thyroid hormone receptor beta (THR-beta).
GN THRB OR NRI42.
OS Parathyroid hormone (PTHrP).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Pleuronectiformes;
OC Pleuronectidae; Parichthyidae; Parichthys.
OX NCBI_TaxID=8255;
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE=96149097; PubMed=8536930;
RA Yamano K., Inui Y.;
RT "cDNA cloning of thyroid hormone receptor beta for the Japanese
RT flounder";
RL Gen. Comp. Endocrinol. 99:197-203(1995).
CC -1- FUNCTION: HIGH AFFINITY RECEPTOR FOR TRIIODOTHYRONINE.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, BETA-1 (SHOWN HERE) AND BETA-2;
CC SEEM TO BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NRI SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: D45245; BAA08201.1; -
DR HSSP: P10828; 2NLL.
DR InterPro: IPR000536; Hormone_rec_11g.
DR InterPro: IPR001723; Steroidhormone_receptor.
DR InterPro: IPR001628; zf-C4.
DR Pfam: PF00104; hormone_rec; 1.
DR Pfam: PF00105; zf-C4; 1.
DR PRINTS: PR00398; STRDHORMONER.
DR PRINTS: PR00047; STROIDPINGER.
DR SMART: SM00430; HOL1; 1.
DR SMART: SM00399; ZNF_C4; 1.
DR PROSITE: PSS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Multigene family; Alternative splicing.
FT DOMAIN 1 31 MODULATING.
FT DNA_BIND 32 99 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 32 52 C4-TYPE.
FT ZN_FING 70 94 C4-TYPE.
FT DOMAIN 149 395 LIGAND-BINDING.
SQ SEQUENCE 395 AA; 45113 MW; F4319CDF96FE2451 CRC64;

Query Match 48.6%; Score 35; DB 1; Length 395;
Best Local Similarity 42.9%; Pred. No. 63;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 GEIQTKPRVQAT 14
: 111:1111
DB 175 GYKEDKPEELGQAS 188

RESULT 15
HYDH_ECOLI STANDARD; PRT; 465 AA.
AC P14377;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Sensor protein hydH (EC 2.7.3.-).
GN HYDH OR B4003.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RX SEQUENCE FROM N.A.
RP STRAIN=K12;
RC MEDLINE=89327164; PubMed=2666400;
RA Stoker K., Reijnders W.N.M., Oltmann L.F., Stoutthamer A.H.;
RT "Initial cloning and sequencing of hydH, an operon homologous to
RT ntrBC and regulating the labile hydrogenase activity in Escherichia
RT coli K-12";
RL J. Bacteriol. 171:4448-4456(1989).
CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM HYDH/HYDG
CC INVOLVED IN THE REGULATION OF THE LABILE HYDROGENASE ACTIVITY.
CC HYDH MAY FUNCTION AS A MEMBRANE-ASSOCIATED PROTEIN KINASE THAT
CC PHOSPHORYLATES HYDG IN RESPONSE TO ENVIRONMENTAL SIGNALS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane